

Stealing a March on Financial Crises: the Scope for Applying Corrective Taxes and Fees towards the Financial Sector – why, what, how

A Review of the Literature

Aleksandar Nikolaev Stoychev



Samfunnsøkonomisk analyse

Department of Economics

University of Oslo

Submitted: October 2014

Table of Contents

Preface.....	v
Executive Summary	vi
1. Introduction.....	1
1.1. Motivation for the current composition	1
1.2. Status of knowledge	2
1.3. A plan for the current composition	2
2. Know Your Enemy: Financial Crises in General	3
3. Financial Crises: an Endogenous Product of the Capitalist Economy	4
3.1. Rationale.....	4
3.2. The role of over-indebtedness	6
3.3. Deflation (of asset prices).....	7
3.4. The credit cycle and the boom-bust/upswing-downswing cycle	7
3.5. Multiplicity of equilibria	8
3.6. Stability of the financial system and destabilizing forces	9
3.7. Memory of financial processes and the economy's development through time	11
3.8. Are financial assets real-world examples of Giffen goods?	12
3.9. Financial innovation	12
3.10. Speculation	13
3.11. Brevity of financial memory	14
3.12. Psychological factors	14
3.13. Role of monetary policy	15
4. The Recent Crisis	15
4.1. Origins of the recent crisis.....	15
4.2. The recent crisis fits into the historical pattern	16
4.3. The recent crisis – a collection of crises	17
4.4. Causes of the recent crisis.....	17
4.5. The recent crisis in Norway	20
5. Containing Systemic Risk – a Task of Macroprudential Policy	21
5.1. Macroprudential policy and macroprudential regulation.....	21
5.2. Systemic risk, financial amplification and procyclicality	22
5.3. Challenges to macroprudential policy.....	23

6. General Considerations – Taxes and Regulatory Command as Tools for Implementing Macroprudential Policy	24
6.1. Differences between taxation and regulation	24
6.2. Common features to taxation and regulation.....	25
6.3. Role for corrective taxes and fees.....	26
7. History of Proposed Corrective Taxes before the Recent Crisis.....	26
8. Financial Sector Externalities	27
8.1. Types of externalities	28
8.2. Other externalities	30
9. Application of Pigouvian Taxation to the Financial Sector – General Considerations	30
9.1. Pigou’s idea	30
9.2. How (dis)similar are environmental and “financial pollution”?.....	30
9.3. How to measure the amount of the externality?.....	31
9.4. What can the various proposals to employ corrective taxation achieve?	32
9.5. What is the implementable optimal corrective tax likely to look like?.....	33
9.6. How is the optimal Pigouvian tax likely to depend on the parameters in the economy?	33
10. Incentives	34
10.1. Systemicness bias.....	35
10.2. Advantages of being too big to fail.....	36
11. Debt Bias.....	36
11.1. Tax distortions and incentives to acquire debt	36
11.2. Policy response.....	37
11.3. The interaction of leverage and risk.....	38
12. Booms and Busts in Credit and Asset Prices	39
13. Information Issues.....	39
13.1. Gains from reducing the scope of information distortion	40
13.2. How could taxes and fees be used to address information issues?.....	40
14. Size of the Financial Sector.....	41
14.1. The size of the financial sector from the perspective of a particular country	41
14.2. Why the financial sector can easily become large in general	42
15. The Bounds of Systemicness	43
16. Types of Proposed Taxes.....	44
17. Financial Transaction Taxes.....	44
17.1. Getting the amounts of liquidity, transactions and speculation right	44

17.2.	Alleged advantages	45
17.3.	Alleged disadvantages	46
17.4.	Theoretical results on FTTs.....	46
17.5.	Empirical results on FTTs	46
17.6.	Effect of FTTs on volume, liquidity and volatility	47
17.7.	Incidence	48
17.8.	Distortions	48
17.9.	Practical issues.....	49
17.10.	Examples of implementation	50
17.11.	Judgement.....	51
18.	Repealing the VAT Exemption of the Financial Sector	52
18.1.	Reasons for exemption.....	52
18.2.	Distortions caused by the VAT exemption	53
18.3.	Difficulties with and potential drawbacks of repeal	53
18.4.	Concrete proposals for rectification.....	54
18.5.	Empirical results	56
18.6.	Alternatives	56
19.	Financial Activities Taxes	57
19.1.	Proposed variants.....	57
19.2.	FATs as a complement to, and an improvement over existing arrangements	60
20.	Taxing Variable Payment.....	61
20.1.	The adverse impact of bonuses on incentives	62
20.2.	The necessity for regulating bonuses.....	63
20.3.	Practical difficulties	63
20.4.	Empirical findings	64
20.5.	Alternatives	64
21.	Bank Taxes	64
21.1.	Concrete proposals.....	66
21.2.	Incidence	67
22.	Stability Fee	68
22.1.	Design	68
22.2.	Issues to take into account when designing the fee	70
22.3.	Incidence	71
22.4.	The proceeds	71

23.	Other Issues	72
23.1.	International coordination	72
23.2.	The adaptability of the financial sector.....	72
23.3.	Do not put too many eggs in a single basket	72
23.4.	Tax havens.....	73
23.5.	Backward-looking taxes.....	73
24.	Conclusion	74
	References.....	77

Preface

I would like to express my gratitude to Vidar Christiansen for being an excellent supervisor throughout the rather long writing process: from helping me find a topic on which my thesis could focus in the very beginning, through suggesting an initial set of references, as well as a few more later on, over discarding an early draft which had derailed completely off course, to reading through the final version of the current composition, with many fruitful and insightful conversations' being scattered along the way. Above all, however, I would like to thank him for his patience with my virtually incessant failing to keep any (well, most) deadlines for submitting the next updated draft.

Executive Summary

As part of the general discussion about the financial sector, its future role and interaction with the rest of the economy, which it spurred, the recent financial crisis has given impetus to theoretical and empirical research aiming to improve the subject of economics' understanding of financial processes, their impact and the management thereof. My focus in the current composition will be mostly on ex-ante issues, more specifically the potential scope for applying corrective (Pigouvian) taxation to the goal of ensuring the financial sector's being more robust and serving its purposes as postulated by economic theory in order that pain from unnecessary turmoil in this area of business be avoided.

For attempts at policy reform to be effective, knowledge about the problem one faces — financial crises — is a pre-requisite. However, mainstream economic theory has so far failed to develop adequate understanding of their nature, a fact reflected in the lack of a generally accepted definition of what a financial crisis is, and of agreement on what causes them. Outside the mainstream there have been put forward theories rationalizing these phenomena, some of which, first and foremost Hyman Minsky's Financial Instability Hypothesis, have received attention and recognition in the post-crisis literature, some of the insights thereof I have found justified in presenting. A common trait of all these non-mainstream contributions is that they explain how financial crises occur endogenously, as a result of the internal workings of the economic system.

To begin with, excessive debt, resulting from financial innovation and what ex post appears as too loose monetary policy, is a key ingredient in virtually any major financial crisis. Further, according to The Financial Instability Hypothesis, a capitalist economy is characterized by a multiplicity of locally stable equilibria, each corresponding to a different mode of behaviour. Equilibrium selection and domain of stability are endogenously determined, the former by institutional arrangements, financial relations and the time path of the economy, the latter by liability structures, more specifically by the mix of three types of borrowers – hedge, speculative, and Ponzi. In addition, mainstream economics literature ignores certain aspects of financial assets, such as their being often purchased for their potential to change price and their behaving as Giffen goods (the term is used nowhere in the reviewed literature, though). Lastly, there are various factors not incorporated in economic models, which nevertheless matter in the real world.

The magnitude and impact of the recent crisis suggest that it was the culmination of a decade-long development. Further, together with its aftermath, it fits well into the historical pattern. Excess liquidity and improperly structured incentives are regarded as the main “contributors” to the recent crisis. These, in turn, are attributed to three factors. First, there are current account imbalances. Second, there is misplaced economic policy, which took form in deficient regulatory framework's failing to oversee systemic risk, use of inadequate models and a policy of low interest rates. Third, there are the results of the changes in the way financial institutions operate and earn their profits: increased interdependence between institutions, increased concentration and complexity making a larger portion of the financial sector systemic, increase in share of profits from fee-based activities, short-termism, increased reliance on leverage, especially short-term.

The wider costs of the recent crisis suggest that there are significant gains to reap from dealing with financial instability. The latter is a task of macroprudential policy, within which the scope for potential use of corrective taxation falls. While the bulk of the post-crisis efforts has been directed towards improved regulatory command, there are contexts within which Pigouvian taxation can achieve or contribute to Pareto-superior allocations, owing to the different properties of the two approaches. The consideration of corrective taxes as a policy instrument is justified by the fact that

the recent crisis was a grossly socially inefficient allocation, which violates The First Main Theorem of Welfare Economics and whenever that is the case, one speaks of market failure; in this instant, the form it took was obvious: the operational framework of the financial sector was conducive to its exerting significant negative externalities on the rest of the economy, which are unified under the common term “systemic risk”.

Several specific realizations of the latter have been identified in the literature. The key ones are pecuniary externalities, to which the acquisition of excessive leverage, heavy reliance on short-term debt, and superoptimal risk-taking can be attributed. Next comes systemic externality which is conducive to too high indebtedness. A third subtype, which goes under many different names, is risk externality, or as I find it most appropriate to refer to as *risk shifting*, which arises because of the presence of explicit or implicit guarantees of support measures towards the financial sector, which, in turn, provide in effect an implicit subsidy to the financial industry, allowing it to borrow cheaper and take more risk. Lastly, there is the failure externality, which again leads to “subsidized borrowing”, but unlike the previous subtype, this one results from the gap between socially optimal and privately optimal levels of capital adequacy. The last two subtypes of externalities enable the superoptimal growth of the financial sector.

The presence of these externalities meant that individual actors in the financial system, who behaved in accordance with theory — they tried to exhaust the opportunities for rent maximization provided by the regulatory framework in place — faced many wrong incentives. Ex post, the most obvious one is that it was not privately optimal to be prudent. Further, it was privately rational for each financial institution to try to become systemic — I call this phenomenon *a systeminess bias*. Next, pre-crisis regulation failed to off-set the incentives to acquire what from a social point of view turned out to be too much debt. I refer to this systematic preference for leverage over other forms of financing, which is conducive to the build-up of eventually unsustainable liability structures, as *a debt bias*. The important interaction of leverage and risk is still awaiting a thorough study. Lastly, there is agreement in the literature that the way the crisis was resolved has increased the scope of moral hazard to which participants in the financial markets are exposed.

Corrective taxes in general are appealing, but problematic, because of, among others, the “knowledge problem” and their potential for inducing undesirable repercussions elsewhere. According to the dominant view in the literature, financial sector externalities possess certain features which present challenges to the direct transfer of solutions used to counter environmental externalities. The difficulties notwithstanding, all issues highlighted by the recent crisis can, at least in theory, be addressed satisfactorily through the use of Pigouvian taxes. In practice, however, currently there is much less that can be achieved. Some issues which they may be used to address include countering the systematic interaction between debt accumulation and asset prices through a counter-cyclical Pigouvian tax on borrowing and the socially superoptimal size of the financial sector. Further, I propose the use of an ex-ante fee in the form of a deposit for information provision as a means of dealing with another form of market failure revealed by the recent crisis — the presence of asymmetric and/or incomplete information. There appears to be broad agreement on certain features of the concrete realizations of corrective taxes: they will be institution- and state-specific, non-linear, most likely variable over time.

All the variants of taxes and fees put forward in the aftermath of the recent crisis, regardless of goal(s) pursued, can be divided into three broad categories: Financial Transaction Taxes (FTTs), Financial Activities Taxes (FATs) and levies on financial institutions’ liabilities.

In the aftermath of the recent crisis, the use of FTTs as means of curbing excessive speculation has been considered, together with other roles they could play, such as discouraging reliance on short-term funding and as a provisional tool to substitute for inadequate regulation in the short run. Empirical results on their claimed benefits and disadvantages are ambiguous. Further, their incidence

is unclear, and they entail various distortions, which may even lead to counter-intended effects. The overall assessment seems to indicate that if employed they will most likely fail to align the behaviour of the actors in the financial system with the socially optimal one, and will, in addition, be welfare-reducing. For these reasons, their use as tools of macroprudential regulation is (currently) virtually unanimously advised against.

One area where use of FTTs nevertheless has been considered is the repeal of the VAT exemption enjoyed by the financial sector. In spite of the body of empirical studies on the welfare effects of implementing such a change's being thin, there is little in the literature to suggest that such a policy is the best way to address the issue.

FATs, on the other hand, are generally viewed favourably in the literature, which can be explained by efficiency reasons, namely they will fall on rents enjoyed by the financial sector, thus appearing to be suitable instruments to employ to achieve the various goals policy makers may have towards financial industry. Further, they are versatile, as there are both instances where they can usefully complement or outperform existing arrangements, generating significant income in the process, as well. Despite their advantages, implementation of FATs still poses certain challenges.

Two potential applications of FATs have found place in the current review. First, taxation of performance-based pay has been put forward as one possible means of addressing the distortion of incentives towards short-termism and excessive risk-taking, which such remuneration arrangements entail. Theoretical and empirical results on the topic are not numerous, and the current body of literature finds no convincing support for the idea. Second, the repeal of the VAT exemption of the financial sector through a kind of FAT is briefly considered.

Two forms of levies on liability structures are presented in the current composition: bank taxes and a stability fee. The former have been considered as an alternative to capital requirements in ensuring the capital adequacy of financial intermediaries. The concrete proposals for implementation more or less agree that the base would consist of uninsured liabilities. There seems to be scope for their being effective. The incidence of the tax will depend on the competition structure of the banking sector.

A stability fee is seen as a viable alternative to capital requirements in their role of aligning the incentives of the financial sector towards decreased acquisition of systemic risk. Various ways of concrete implementation have been proposed. The stability fee can be the source of significant income, which can be used to fill a resolution fund, for instance.

1. Introduction

1.1. Motivation for the current composition

The financial crisis in the autumn of 2008 turned the world around. While victory over it was declared within months, its repercussions are still felt. Among (many) other things, it provided fertile ground for research to both theoreticians and practitioners of the economics profession. Its impact has spurred a discussion about the financial sector, its role and interaction with the rest of the economy in the future. The crisis may well become the source of forces which will lead to a paradigmatic shift away from the neo-liberal dogma, the essence of which is suitably described by the yet-to-be-accepted term “financialization”¹. More generally, one should expect the economics profession to develop better understanding of financial processes, especially from a holistic perspective, so that it can provide solid foundation for effective and efficient policy actions. This translates in centring attention on the ex-ante and ex-post aspects of developments in the financial sector in general and financial crises in particular.

In the current composition I will focus mostly on ex-ante issues, which have to do with what I term “prevention and mitigation of crises”. This might sound rather boring, especially in our days, when there seems to be agreement that not a single day of our lives should be allowed to pass into history without drama. The recent crisis, however, showed the costs of suspending precaution and *re-acting* to events, especially in an ad hoc manner². I espouse a *pro-active* view: one should be focused on preventing calamities from happening, rather than just waiting for them to occur and then play hero trying to deal with them. Unfortunately, the trend has been to move decisively away from this view in the last few decades; it has been rendered obsolete, to use a neutral word. Still, if one insists on playing hero, one should at least do it properly. Within the context of the post-crisis reform agenda, this has been recognized and found expression in efforts to create effective resolution mechanisms for failing institutions; however, these are mostly ex-post issues, so they will not receive but marginal attention in the current composition.

Within ex-ante issues I will look more specifically at the potential role corrective taxation could play in the attempts to reduce the probability of the inevitable crises, and to avoid the unnecessary ones. There are several reasons for why I have chosen this topic. First, according the supervisor of the current composition’s author, this subfield is still largely undeveloped³, so there is a lot of “unchartered territory” to explore. Second, I happened to study economics when the recent crisis struck. Thus, I have had the opportunity to witness both the time when the profession was at its height, marked more or less clearly by the neo-liberal agenda and perception of the world, and the disarray which followed after the aforementioned watershed. In the post-crisis period economics seems to be characterized by a dichotomy: on the one hand, there is insistence on the continued reliance on a paradigm refuted by real-world events, on the other, there is intensified interest in and research of previously disregarded or dismissed ideas. As part of my research for the current writing, I have come across several such works and it surprised me that they often rationalized events the existence of which mainstream economics does not even allow. For these reasons, it is to be hoped that the recent crisis will lead to a much-needed discussion about an update and upgrade of

¹ One definition of the phenomenon can be found in Epstein (2005:3): “...financialization means the increasing role of financial motives, financial markets, financial actors and financial institutions in the operation of the domestic and international economies.”

² There is agreement in the literature that the mess which the crisis engendered is attributable to the lack of mechanisms for quick, orderly “resolution” (another neologism for an old thing — liquidation — that came into use as a result of the recent episode of distress, besides “quantity easing” — the new way to refer to money printing) of financial institutions in trouble.

³ This does not mean that the idea is a new one: already Keynes in his *General Theory* (1936) has considered it.

economists' toolbox. This composition is an attempt to contribute to this goal. Third, it is interesting to see whether Pareto-improved allocations are achievable through the use of the market mechanism. Besides its appeal to free-marketeers, it is worth considering what part it can play in the new regulatory framework, which aims at a more robust financial sector. This is especially important in view of the fact that a radical shift towards an arrangement akin to the one in existence in the two decades following the Second World War does not seem very likely in the foreseeable future.

1.2. Status of knowledge

It is meaningless to speak of the status of knowledge on a particular subject without a reference point in time which serves as the limit of the surveyed period. As was said above, the recent crisis shook profoundly the economics profession. Therefore, it seems natural to use the strained episode, which culminated with the trouble in the autumn of 2008, as a reference point for the purpose of the current composition. However, nothing happens in an instant, so I think it is justified to set the reference point as of 31st December 2008. By that time the crisis was more or less vanquished, but its impact on the wider economy had just begun to be felt. At around the same time (most) theoreticians and practitioners of the economics profession had finally acknowledged that something extraordinary had happened. It provided ample opportunity for research and studies. Indeed, the bulk of the resulting writings, whatever their official type, started appearing in 2009.

So, what did we know by the 31st December 2008 about the potential role of corrective taxation as a tool of regulation of the financial markets? The literature, which I have reviewed for the purpose of the current composition, is unanimous that both on the theoretical and the empirical field there was virtually nothing extant. In his *General Theory* (1936) Keynes did propose the idea to use taxes as a means of curbing the speculative excesses in the financial markets. No development of his idea followed, however. This may be explained by the lack of such a necessity: the economic framework put into place after the Second World War solved the issue until about 1970. The next proposal came as a possible remedy to the disarray which followed the collapse of the Bretton Woods arrangements: the Tobin tax, as it is known, focuses too narrowly on the foreign exchange market, though. Further, there is no recognition that the financial sector exerts significant negative externalities on the rest of the economy. In general, it may be argued that the negative effects of the development towards "financialization" and the necessity to counter them did not receive any significant attention until after the recent crisis. This is without doubt one of the reasons for the lack of any significant body of literature on the topic of applying corrective taxation to the financial sector.

1.3. A plan for the current composition

I will begin the survey of the literature with a brief description of financial crises. It will be an attempt to fill the void left by standard theory when it comes to rationalizing their occurrence and impact. Understanding the nature of financial crises is an obvious precondition for taking a pro-active stance; lack of comprehension of events of distress in the financial markets will render any measures implemented ineffective, and even counter-productive. Next I will look at the various issues revealed by the recent crisis and the scope for using corrective taxation in addressing them. My focus here will be on the question whether taxation achieves the goal of changing the behaviour of the actors in the financial markets in a socially favourable direction, thereby usefully supplementing direct regulation, or outperforming it, wherever possible. Lastly, I will consider some concrete proposals which have been put forward since 2009. Virtually all proposed and implemented taxes have had also other motives, such as revenue-raising and retribution. To the extent possible, these other motives will be ignored, unless they are essential for the setting in question. In addition, while my focus is on ex-ante issues, ex-post concerns will be taken into account, where this has been identified as important.

2. Know Your Enemy: Financial Crises in General

One of the goals on the post-crisis agenda has been the development of a robust financial sector. In order to achieve this one has to be aware of the problems one faces. Knowledge of the nature of financial crises is therefore indispensable. The contribution of modern mainstream economic theory to our understanding of these phenomena has been modest. This can be one of the explanations behind the standard reaction to a crisis, whenever one occurs: a mixture of surprise, disbelief, confusion.

To begin with, we need a definition of a financial crisis. In spite of the fact that events termed “financial crises” have occurred repeatedly for at least four hundred years, there is no generally accepted definition of a financial crisis, except that it is something which entails large disturbances in the financial markets and the financial sector. In short, one knows a financial crisis when one sees it. The following working definition, inspired by the 1987 article “The Anatomy of Financial Crises” by Eichengreen and Portes, can be found in NOU (2011:1, p.31):

[A] financial crisis is a powerful disturbance in the financial markets, typically associated with strongly falling asset prices and insolvency among borrowers and financial enterprises, which spreads throughout the financial system, disturbs the market’s ability to function and gives significant effects on activity and employment. In an international financial crisis the disturbance spreads across country boundaries and disturbs the markets’ ability to function internationally.⁴

While such a definition based on qualitative indicators may be useful for theoretical pondering, for practical purposes the reliance on quantitative indicators is indispensable. One example of application of the latter can be found in Laeven and Valencia (2010b), who in their definition of systemic banking crises rely on both qualitative and quantitative criteria being met at the same time, where the latter are based on the figures of implemented support measures towards the banking sector. While the focus of the two authors’ work is on systemic banking crises, the extension of their results to apply to financial crises seems straightforward, especially in view of the blurring between banks and non-banks which happened before the recent episode of distress.

The recurrent nature of financial crises has provided ample opportunities for the accumulation of some knowledge about these phenomena. Mainstream economic theory has made limited use of it, however.

Paralleling natural disasters which occur at irregular intervals, financial crises share some common features, while each is different. Unlike natural disasters, however, there is no agreement on what causes financial crises. This uncertainty about the causes of these phenomena is the main reason why their prediction is very difficult, at least if one relies solely on mainstream theory. While it is possible to observe the development of a crisis, for example by using deviation from trends in various indicators, no one within the mainstream has so far been able to use this and other information to pinpoint the time a crisis will hit and how large its impact will be. This fact is also a challenge to policy, as acknowledged by Minsky (1986) and NOU (2011:1): when measures are designed to promote financial stability, it is important to keep in mind that the next crisis almost certainly will be quite different in nature.

⁴ This is a very literal translation of the Norwegian text:

En finanskriser er en kraftig uro i finansmarkedene, typisk forbundet med sterkt fallende aktivapriser og insolvens hos lånetakere og finansforetak, som sprer seg gjennom det finansielle systemet, forstyrrer markedets funksjonsevne og gir betydelige utslag i aktivitet og sysselsetting. I en internasjonal finanskriser sprer uroen seg over landegrensene og forstyrrer markedenes funksjonsevne internasjonalt. (NOU 2011:1, p. 31)

Financial crises can be triggered by exogenous “shocks” or occur endogenously, as a result of the operation of the financial system and the wider economy. Modern mainstream economics explains such episodes of trouble by the former, while various theories outside it have rationalized them using the latter.

Financial crises tend to occur in the aftermath of deregulations in the financial sector. The latter go almost invariably hand-in-hand with a wave of innovations or “innovations”, nowadays termed “financial engineering”. High international capital mobility is often a precondition for financial (especially bank) crises. The risk of financial crises is increased, whether the flow of financial capital is into or out of the country in question, because the large amount of capital enables the build-up of significant imbalances. The latter are always present in a run-up to a crisis. Further, the role of non-banking institutions, nowadays referred to as “the shadow banking system”, increases as a result of their facing less strict regulation. The expansions prior to financial crises have tended to be stronger than expansions prior to ordinary business cycle setbacks. Debt, housing prices, consumption, and employment increase by more than they do in an average expansion. Correspondingly, economic setbacks in the aftermath of financial crises are deeper and of longer duration than other ones. In NOU (2011:1) one finds that precipitous declines in real estate prices often go hand-in-hand with large economic setbacks, while the connection between precipitous declines in share prices and economic setbacks is less pronounced. Further, even a robust financial system can be brought to a crisis by a sufficiently powerful economic disturbance.

Hyman Minsky accorded due attention to the effect of good times, especially on the undervaluation of risk. At least one of the surveys of the recent crisis reiterates his ideas (see, for instance, the first paragraph of section 11.2.1 in NOU (2011:1, pp.: 133-134)). A common trait of the run-ups to crises is that speculation and over-optimism about the future become the dominant sentiment in the market. This affects expectations which are extrapolated and creates a momentum which drives asset prices up. This positive feedback loop operates until the imbalances, to which the entire process leads, can no longer be sustained.

One can distinguish between various types of financial crises according to where in the financial markets they occur and how they develop.

One last, but not less unpleasant feature of financial crises is that, for the foreseeable future, they are sure to keep coming. This is contended by NOU (2011:1), Kocherlakota (2010), Minsky (1982; 1986), Galbraith (1993; 1995) and Cooper (2008).

Financial (and other) crises benefit the subject of economics, because they spur discussions and research, which often improve the discipline’s understanding of the way the economy behaves. In fact, it has been acknowledged in the literature that crises and episodes of trouble often are discipline-defining and discipline-changing events.

3. Financial Crises: an Endogenous Product of the Capitalist Economy

3.1. Rationale

One of the things about the recent crisis, which appeared strange to me, was that it astonished more or less everybody versed in the subject of economics, and especially so the ones deemed, or as it is more precise to say, revered as experts⁵. It was even weirder that this turned out to be a not-one-

⁵ And particularly so the most vocal and media-exposed ones among them (“most vocal and media-exposed” is the modern equivalent of what was known as “prominent” in the old days)

time event: Galbraith (1993) explicitly points out the fact that the aftermaths of speculative episodes always evoke surprise, wonder and enthusiasm. This suggests that not only is an important feature of our type of economy not understood, but it is even ignored. This hides dangers, against which already Minsky in *Stabilizing an Unstable Economy* (1986) has warned. More specifically, he argues that if theory is at variance with the way the economy behaves, reforms will do little good and may potentially do a lot of harm:

Before we can do better, we must understand our economy. Unfortunately, policy makers and advisors are the slaves of an economic theory that misspecifies the nature of our economy by ignoring instability. That perhaps is a true measure of our crisis: nobody “up there” understands American capitalism. (Minsky 1982:202)

Instability is an observed characteristic of our economy. For a theory to be useful as a guide to policy for the control of instability, the theory must show how instability is generated. ...Once instability is understood as a theoretical possibility, then we are in a position to design appropriate interventions to constrain it. (Minsky 1982:xii)

He identifies the inability to explain financial crises and financial instability as the major shortcomings of conventional theory. Some three decades later and with dozens of additional Nobel laureates, these challenges posed by reality to economic theory still await their satisfactory resolution.

There are, however, theories, not incorporated in the mainstream canon of the economics profession, which deal with these issues. The recent crisis is a good opportunity to rethink their merits, a point made in Bean (2009)⁶ and NOU (2011:1). I will present relevant aspects of some of these theories. The ones reviewed include Irving Fisher’s “Debt-Deflation Theory of Great Depressions”, Hyman Minsky’s Financial Instability Hypothesis, the Credit Cycle of The Theory of Long (Kondratiev) Waves⁷, the Austrian School’s view of loose monetary policy’s being the factor behind the credit cycle, which in turn drives business cycles. Their place here have also found some relevant (and witty) observations by John Kenneth Galbraith and a book by George Cooper, published while the recent crisis was still a liquidity problem. I have made an attempt to synthesize the various aspects these theories have, rather than presenting them one after the other. My main criterion for selection was the appropriateness of each of the insights provided by these theories for bridging the gap between theoretical predictions and empirical realizations.

The endogenous nature of financial crises (and their precursors the financial excesses) has been asserted by various authors:

Recurrent speculative insanity and the associated financial deprivation and larger devastation are, I am persuaded, *inherent in the system* [emphasis added]. Perhaps it is better that this be recognized and accepted. ... the things that signal their *certain* return [emphasis added]... this process, once it is recognized, is clearly evident, and *especially so after the fact* [emphasis added] (Galbraith 1993:viii)

The hypothesis is that the financial panic which is present during deep depressions and absent during mild depressions is not a random exogenous affair; rather *it is endogenous to the economy* [emphasis added]. The financial panic is made possible by the changes in the

⁶ As a matter of fact, I was made aware of the work of Hyman Minsky (and himself) by Bean’s speech.

⁷ Unlike other economic cycles (Kitchin, Juglar and Kuznets), the existence of Kondratiev waves is not accepted by mainstream economic theory. The contentious nature of the theory should not be used as an argument against its inclusion in this part of the current review. What matters is the ability to match reality, and explain it plausibly, too. The inclusion of the Long Wave Theory should not be considered as an attempt of my arguing in its support — such a question is obviously outside the scope of the current composition; rather, it would merit a separate thesis.

financial structure which take place during the long swing expansion. ((Minsky 1964:325), *Longer Waves in Financial Relations: Financial Factors in the More Severe Depressions* quoted in Tylecote (1992:97))

In asserting the existence of long waves and in denying that they arise out of random causes, we are also of the opinion that *long waves arise out of causes which are inherent in the essence of the capitalist economy* [emphasis added]. (Mager 1987:32)

He [Kondratiev] makes the point that cycles were not isolated factors in history but are part of connected elements in technological progress and political and social history, and are based on *endogenous factors*" [emphasis added]. (Mager 1987:41)

The most comprehensive theory among the ones I have reviewed is Minsky's Financial Instability Hypothesis, which is also his main contribution to the economics profession, in spite of the fact that its merits have yet to be granted due recognition by the latter. It builds on an interpretation of the ideas of Keynes (and further on Fisher and Simons, see Minsky (1986:172)). Minsky (1982) argues that some of the crucial insights of Keynes were lost, when only part of his work was incorporated in standard economic theory. One such instance is the operation of forces of disequilibrium in the financial markets, which directly affect the valuation of capital assets and for this reason are important in the determination of investment activity. Another one is that money enters the economy "in an essential and peculiar manner" (*The General Theory* (1936:vii), quoted in Minsky (1982)). Minsky's main result is that financial and, more generally, economic crises can and do arise endogenously as a result of the functioning of the modern capitalist economy. He identifies three features of the latter which allow the occurrence of such phenomena: privately owned capital assets, complex and constantly evolving financial institutions and practices, and the fact that the economy develops through time and has a history.

3.2. The role of over-indebtedness

Fisher, Minsky and the proponents of the Credit Cycle Theory within The Long Wave Theory identify excessive debt as more or less *the* economic variable capable of causing the great disturbances in the economy which produce great booms and depressions. Building on Minsky (1982; 1986), I define "over-indebtedness" as the emergence over time of liability structures which can be sustained only if euphoric expectations are fulfilled. Euphoria, discussed in Minsky (1982; 1986), Galbraith (1993), and Mager (1987), is simply the excessive speculation which each prolonged boom climaxed with. Fisher, on the other hand, sees over-indebtedness as a multidimensional magnitude, a relative quantity to other items. This means that it is not given by the mere amount of currency units owed, but other features should be taken into account as well; for instance, he points out the risks of... short-term borrowing.

How does over-indebtedness arise? Fisher argues that it is the result of over-investment. He identifies several "debt starters", of which the most common one is "new opportunities to invest at a big prospective profit" (Fisher 1933:348). In a similar vein, the proponents of The Long Wave Theory argue that the debt burden, which starts biting in the downswing, is accumulated during the early stage of each new wave, when there is high investment to support the diffusion of new technologies. To Fisher, another "debt starter" is low-interest-rate policy, because it provides cheap liquidity to fund these endeavours with: "[e]asy money is the great cause of overborrowing" (Fisher 1933:348). On the other hand, to the adherents of the Austrian School, low-interest-rate policy is *the* "debt starter". Fisher does not believe that self-confidence is to blame, except when it "beguiles its victims into debt" (Fisher 1933:341).

One final aspect of over-indebtedness is what can be called “declining marginal productivity of debt”⁸. Within The Long Wave Theory, it describes the combination of increasing debts and the declining effect they have on economic expansions, while Minsky (1982; 1986) and Cooper (2008) refer to the gradual loss of ability of using debt stimuli to deal with recessions. The topic and some of the difficulties governments face in our days are explored in some detail in NOU (2011:1).

3.3. Deflation (of asset prices)

Irving Fisher identifies deflation as the second major factor behind the excessive booms and depressions observed throughout history. Within the context of financial crises, it is asset price deflation which is important; general decline of other prices is a consequence of the reduced economic activity which ensues in the aftermath of the disaster in the financial sector. Deflation has been incorporated in the model of financial amplification, formalized in Bernanke, Gertler and Gilchrist (1999), but already Fisher was aware of its most devastating feature:

Then, the very effort of individuals to lessen their burden of debts increases it, because of the mass effect of the stampede to liquidate in swelling each dollar owed. Then we have the great paradox, which, I submit, is the chief secret of most, if not all, great depressions: The more the debtors pay, the more they owe. (Fisher 1933:344)

The arrest of this self-reinforcing process has been the officially stated rationale for interventions in support of troubled institutions. Minsky refers to assets the price of which will be stabilized as “protected assets”⁹. He argues that the ratio of unprotected to protected assets is one of the major factors which determine the stability of the financial system. A drawback of these protected assets, which Minsky points out, is their effect on incentives: every time the central bank (or the government, for that matter) protects a financial instrument, it legitimizes its use as a source of finance — in modern terminology we would say that this increases the scope for moral hazard.

According to the adherents of the Austrian School and the proponents of the Long Wave Theory, deflation is a necessary and unavoidable correction mechanism, when markets clear the amassed real imbalances, which have been caused by nominal imbalances. This means that attempts by monetary and fiscal authorities to prevent deflation by sustaining asset prices and profits (through demand stimulation) cannot be sustained over the long run¹⁰.

3.4. The credit cycle and the boom-bust/upswing-downswing cycle

The credit cycle is the main driving force behind the boom-bust cycles according to the adherents of the Austrian School and, together with innovations, behind the dynamics of the long waves. It is defined as the expansion and contraction of the access to credit and its amount over the course of the business cycle, or the long wave. To the adherents of the Austrian School, the central element in the credit cycle is what they regard as poorly conducted monetary policy: central banks have the habit of engaging in excessive monetary expansion over extensive periods which results in too much

⁸ According to the Wikipedia article “Kondratiev Wave”, the concept of “marginal productivity of debt” was developed by Melchior Palyi (1892-1970) to describe the ratio of GDP increase for a unit’s increase in a nation’s debt, although the term itself came into use after his death.

⁹ The stabilizing is formally carried out by the fiscal authorities, the central bank, or both, in the form purchases (for example bonds and savings deposits) or loans. The residual payer in any case is invariably the taxpayer, though.

¹⁰ There are at least two other reasons for the unsustainability of such a policy. First, it would require that all agents suffer from money illusion — otherwise they would not accept the engineered inflation — but mainstream economic theory assumes that they do not. Empirical studies have confirmed the existence of money illusion, but the question is whether it can be sustained over the long run. Second, the declining marginal productivity of debt combined with already high liabilities of fiscal and monetary authorities act as a limiting factor to their ability to arrest a major price decline.

debt's being created, which in turn is used to feed speculative bubbles. Within the Long Wave Theory, it is the necessity to finance the diffusion of new technologies and the variation in their profitability which cause fluctuations in the availability of credit. On the other hand, Minsky identifies financial innovation as the source of abundant credit.

How do the variations of the credit cycle translate into series of booms and busts? The boom phase starts when there is increased liquidity, whether because of monetary expansion (Austrian School) or because of financial innovation (Minsky)¹¹. The borrowing increases demand and investment. This results in increased capital gains, which, together with increased spending increases profits. The pursuit of more profits raises the price of capital assets and therefore the demand price of investment, which combined with the abundance of credit fuels asset price inflation, driven by frenzy to reap any remaining, yet-unused, or new investment opportunities. With limited financial capital, additional asset acquisitions will be more and more leveraged, with the result that the pace of increase in debt levels will pick up ever more quickly. This will culminate in investing more in certain ventures than the return on the margin justifies. This diversion of capital to socially inefficient uses can take various forms: formation of a bubble or superoptimal increase in consumption among others. All the time the development is further exacerbated by the increased number of people qualified for obtaining a loan, declining lending standards and more financial innovation. This whole process causes the economy to expand beyond any tranquil full-employment state. The financial structure gradually becomes more conducive to a financial crisis. Thus, according to Minsky, a full-employment equilibrium will lead to an expansion of debt-financing, which will move the economy away from it. This transitory nature of an otherwise efficient allocation is due to the investment boom which is brought about by speculation upon and experimentation with liability structures and novel/"novel" financial assets, or artificial stimuli.

The bust part of the cycle commences when the pool of borrowers willing or able to purchase at the inflated prices is exhausted. The prices in (the segment of) the market where most of the excess debt has been amassed collapse triggering a wave of insolvencies and bankruptcies. If a sufficient "critical mass" of the ultimate lenders (the banks) finds their solvency and/or profitability in jeopardy, lending is curbed to avoid further damage. Eventually this results in a credit crunch, in which there is precipitous reduction in the volume of liquidity available. This point at which the credit cycle suddenly turns from expansion to contraction is known in the literature as the *Minsky moment*.

The bust is normally "heralded" by a financial crisis, which can be triggered by a decline of income, or the distress of particular unit, usually a financial one. The triggering event also causes a reversal in public confidence, so that even if it seems insignificant in and of itself, it is the fact that it happens amidst an economy full of imbalances — high debt, low productivity and with inefficiencies in production — which is the reason for its systemic impact.

Examples of such boom-bust cycles only in the past 30 years are provided by the Japanese asset price bubble, the dot-com bubble and the U.S. housing bubble, which triggered the recent crisis.

What was said above is true for the Theory of Long Waves as well, except that it uses a larger timescale and correspondingly the more appropriate terms "upswing" and "downswing".

3.5. Multiplicity of equilibria

The recent crisis reminded us that the economy may operate in several equilibria which differ in their stability and Pareto characteristics. It will hopefully intensify the efforts to understand this multiplicity and the results will be incorporated in mainstream economic theory. Others have already acknowledged its importance: from Fisher who opened for the possibility of unstable equilibria: "The more the economic boat tips, the more it tends to tip. It is not tending to right itself, but capsizing."

¹¹ In both cases liquidity is increased further by the effect of the money multiplier.

(Fisher 1933:344), to the more recent claim by Cooper (2008) that the current financial system is inherently unstable, has no steady state equilibrium and is prone to regularly forming damaging boom-bust cycles.

The multiplicity and stability of equilibria are central to Minsky's Financial Instability Hypothesis. Its essence is what he calls "the big theorem": "...capitalist economy with sophisticated financial institutions is capable of a number of modes of behavior and the mode that actually rules at any time depends upon institutional relations, the structure of financial linkages, and the history of the economy" (Minsky 1982:92). Minsky (1986:173) is more specific by stating what he sees as the two fundamental propositions of the Financial Instability Hypothesis. First, capitalist market mechanisms cannot lead to a sustained, stable-price, full-employment equilibrium in a capitalist economy which welcomes financial innovations, because within any full-employment situation there are endogenous disequilibrating forces at work, which will disrupt tranquillity. Second, serious business cycles are due to financial attributes that are essential to capitalism.

3.6. Stability of the financial system and destabilizing forces

An important feature of the Financial Instability Hypothesis is that it is derived for a system which is not globally stable, but has several locally stable equilibria. What determines the region of stability around each of them? The main argument of "Financial Instability Revisited: The Economics of Disaster" (Minsky 1982:117-161) is that the domain of stability of the financial system is determined mainly endogenously by liability structures and institutional arrangements; the exogenous determinants are government and central banking arrangements. Minsky argues that how large a shock the financial system is able to absorb and still return to its initial equilibrium depends on the financial structure and real income.

He further argues that the fundamental instability of a capitalist economy, observed repeatedly in practice, is upward: there is a tendency to transform good performance into a speculative investment boom. There are two reasons for this: the complex set of market relations which enter into the investment process, and the way the liability structure commits cash flows from producing and distributing output. Nevertheless, Minsky notes explicitly that not all financial structures are equally prone to financial instability: the two decades after the Second World War provide a vivid, though a rather isolated example.

Minsky argues that the financing of investment and positions in capital assets by debt is a distinguishing feature of the capitalist economy. Further, the financial dimension of a *corporate* capitalist economy is much greater than for an economy dominated by partnerships and proprietorships. Hence the central determinant of the stability of the economy are gross profits net of taxes, but including interest payments, because this is what is available to service debt commitments with. With the financing of capital asset ownership and investment as the critical destabilizing phenomenon, it is clear that restrictions on the liability structures available are necessary, if instability is to be contained.

So how are investment and positions in financial assets financed? Minsky divides the financial postures of firms, households and governments into three categories, depending on the relation between the payment commitments to meet their liabilities and their primary cash flows: hedge, speculative and Ponzi. What characterizes each type? For hedge financing units the cash flows from income are expected to exceed contractual payments on outstanding debts in every period. For speculative financing units the total expected income cash flows over the foreseeable future exceed the total cash payments on outstanding debt, but in the short term the payment commitments exceed the income cash flows. Still, net income over the short term exceeds interest payments over the short term. A Ponzi finance unit is a speculative financing unit which in the short run does not have enough income cash flows to even meet interest payments due; hence its outstanding debt will grow. With respect to the recent crisis, we have the following examples of each of the three types:

hedge financing units were most of what is known as the real economy, speculative financing units were the overwhelming majority of participants in the financial sector, and Ponzi financing units were the so called “subprime borrowers”.

Before we continue, it is necessary to specify what is meant by “speculation” in this context. To Minsky, the “speculation” of the speculative units is that refinancing will be available when needed. Speculative units finance long positions in assets by short-term liabilities: commercial banks are a prime example (due to the familiar asset-liability mismatch). Ponzi units engage in speculation in the sense used in the everyday language: they are dependent on (asset price) inflation¹².

What distinguishes speculative and Ponzi financing units from the hedge financing units, on the one hand, and what are the differences between them, on the other? First, both speculative and Ponzi financing units must either borrow or sell (some of) their assets to obtain the means to meet their commitments. The former must borrow an amount which is smaller than the maturing debt, whilst the latter must increase their outstanding debt. If Ponzi units are to obtain finance, their total expected cash inflows must exceed their cash outflows, which means that Ponzi units are dependent on the belief that some of their assets will be sold at a higher price in the future. Second, changes in interest rates cannot affect the solvency of hedge financing units, because the present value of their capital assets exceeds the book value of their debts. Thus, they are not *directly* susceptible to adverse changes in the financial markets. Hedge financing units can only go bankrupt if their revenue falls to the point where they can no longer meet their operational costs and debt commitments. On the other hand, the solvency of both speculative and Ponzi financing units is affected by disturbances to income *and* changes in the interest rates. Thus, unlike hedge units, speculative units can suffer from a present value reversal at high interest rates. Further, the ability of speculative financing units to meet their obligations is *directly* susceptible to failures in the markets where they sell their debts: an example in the recent crisis was the freeze of the interbank market.

The type of a given unit is not necessarily constant through time: a decline in the expected gross capital income can turn hedge financing units into speculative ones and speculative financing units into Ponzi ones. On the other hand, an increase in income or, what Minsky terms “funding” of debts (i.e. bankruptcy) can transform speculative financing units into hedge financing units, and Ponzi financing units into speculative financing units.

Three major factors determine the stability of the financial system. First, it is the adequacy of income cash flows to meet debt obligations. Second, it is the adequacy of refinancing possibilities for debtors who have suffered a drop in income flows to continue servicing their debts at an acceptable price (or at all). Third, it is the ratio of unprotected to protected assets, mentioned before. All three imply that the stability of the economy’s financial structure depends on the mix of the three types of borrowers: the higher the proportion of hedge units, the more stable the economy, whereas an increase in the weight of either speculative or Ponzi financing, or both, increases the susceptibility to financial instability.

Among the determinants of the stability of the financial system of lesser importance is the layering of financial obligations¹³. When it increases rapidly, the domain of stability of the financial system is decreased. The reason is that the increased interconnectedness in the financial sector opens up for more, and more likely potential problems if one unit is unable (for whatever reasons) to meet its commitments, even only temporarily. In other words, the financial system is very susceptible to a

¹² To be precise, speculation in the everyday use of the word refers to betting on price changes in either direction; for instance, when one plays “short”, one hopes for a price decline.

¹³ Unlike Minsky’s days, it has received a lot of attention in the aftermath of the recent crisis. In the post-crisis literature it is identified as one of the central culprits for the fragile financial structure which arose in the years before 2008. It can, therefore, be argued that its importance as a factor behind financial stability has increased.

disruption in the flow of payments. Further, greater financial layering implies higher prices of assets, because even if the additional layers did not operate for profit, they still need to recover the costs for their operation. Thus, the growth of the financial sector is a factor behind asset price inflation. From this it follows that if the size of the financial sector is reduced, so would likely be asset price inflation. With fewer layers, there would be less instability. One way this can be achieved is through taxes, because they would make additional layering unprofitable — this latter point is made in the post-crisis literature.

An important distinction from the standard theory is the claim that tranquillity and success are not self-sustaining states: no external shock is necessary to disrupt them. In fact, Minsky argues that the domain of stability of the financial system will decrease during an extended boom. The reason is that whereas standard theory always rationalizes a boom as a reaction to changes in fundamentals, Minsky points out the fact that history has taught us (or, to be precise, should have taught us) that, if unchecked, a boom inevitably turns into a speculative one. He defines the latter as a situation where a substantial and growing portion of outstanding payment commitments can be fulfilled only if an appreciation of asset values takes place¹⁴. Because the speculative boom is a self-reinforcing process, the proportion of the value and type of financial assets, which are sensitive to sharp revaluations of expectations, will gradually increase, so that over time liability structures emerge which can be serviced only so long the boom continues.

Minsky identifies two facets to financial instability: the cost of debt and the need to roll over ever larger amounts of debt lead asset values' breaking as units try to reduce their dependence on debt, and a fall in gross capital income, because the determinants of profits themselves have fallen¹⁵.

Cooper (2008), who has built on Minsky's work, identifies another two internally-generated destabilizing forces: supply may drive demand in asset markets, and changes in asset prices as a factor behind asset demand. These take various shapes through bank credit creation, mark-to-market accounting, debt-financed asset markets, cyclical dependence of credit spreads, demand driven by scarcity and price. He believes that the key channel through which financial instability can be and is generated is the conflict between guaranteeing some return on capital, while putting (some of) this very same capital at risk.

3.7. Memory of financial processes and the economy's development through time

As mentioned at the very end of subsection 3.1., Minsky identifies the fact that the economy develops through time and has a history, which affects the present, as one of the three features of the modern capitalist economy which mainstream economic theory has failed to incorporate and hence some of its results do not hold in practice. He argues that inclusion of time is one of the necessary extensions of economic theory. Minsky argues that the evolution of the economy through time is the reason why once-and-for-all solutions to its problems do not exist.

More specifically, the time path of the economy depends on the financial structure: the relative amount of debt and equity financing used by firms reflect the cyclical past of the economy. However, the financial structure is itself subject to change as a result of the internal reactions to the success of the economy. The major factor which determines its shape in a capitalist economy is the way investment and positions in capital assets are financed: at any date the performance of the economy is closely related to the success of debtors in fulfilling their commitments and the views on the ability of new borrowers to fulfil theirs, at that point in time. The former will depend on whether the profits,

¹⁴ The housing boom is a recent example, where the hope or belief was that "just this once" housing prices would continue going up forever.

¹⁵ For instance because of market saturation, or the exhaustion of a technological process towards the end of a long wave in Kondratiev's theory.

which arise from cash flows, suffice to validate any given structure of business debt, or do not. On the other hand, what debt structure firms, households and their creditors will consider acceptable is determined by the expected level and the stability of profits.

The proponents of The Long Wave Theory also espouse the view that the development of the economy through time matters:

Thus, each cycle is different because each phase is the “consequence of conditions amassed during the preceding time *internally* [emphasis added]... Each new cycle takes place under *new concrete historical conditions* [emphasis added], of a new level of the development of productive forces, and hence is by no means a simple repletion of the preceding cycle. (Mager 1987:32)

The existence of positive feedback processes requires that history influence current and future events, which means that they must exhibit some memory. Cooper (2008) builds his argument against the reliance on the Efficient Market Hypothesis on this fact. He is further of the opinion that memory of the markets can help explain the fat tails problem and the systematic underestimation of financial market risk¹⁶.

3.8. Are financial assets real-world examples of Giffen goods?

In *Stabilizing an Unstable Economy* (1986) Minsky calls our attention to the observed fact that conjectural and speculative elements in the markets where financial instruments and capital assets are traded often cause these to behave as what we know from theory as Giffen goods (he does not use the term, but the description he gives fits the definition we find in textbooks): in a boom their price increases and so does the quantity demanded; conversely, in a decline, the demand for assets can fall even though their price also has fallen.

Cooper (2008) elaborates further on this observation by Minsky, but neither he uses the term “Giffen goods”. He argues that the mechanism which operates in the markets for assets is fundamentally different from the one in the markets for goods and services. The crucial difference between them is the way they respond to shifts in prices and in demand. In goods markets, higher prices trigger lower demand (and vice versa), in asset markets, higher prices trigger higher demand (and vice versa). Thus, it appears that the former markets are dominated by normal goods, while the latter are dominated by Giffen goods.

3.9. Financial innovation

Financial innovation, referred to as “financial engineering” in the context of the recent crisis, has received attention in the works of Minsky, Galbraith and the proponents of The Long Wave Theory. All of them argue that it provides the liquidity which is used to incur the eventually unsustainable debts. Because of this, it is a major factor contributing to instability.

Minsky argues that financial innovation is more common during good times¹⁷, because the pressure for profits creates demand for additional liquidity. Financial innovation invariably manages to satisfy this demand, because, as noted in the post-crisis literature, there are no serious capacity constraints

¹⁶ This is just one aspect of the problems which use of improper models can cause. The modelling issue has received some attention in the aftermath of the crisis, but it has so far failed to incorporate insights from the pre-crisis literature. At the extreme, the representatives of the Austrian School in general disapprove of the use of models and statistics, because, as they argue, the uniqueness of economic events renders impossible any attempt to describe them probabilistically in a satisfactory way. Minsky (1982, 1986) has also claimed that the uncertainty in the economy cannot be modelled in the same way as probability in games of fortune. Cooper (2008) argues that the entire probability distribution often changes at the Minsky moment.

¹⁷ In *Stabilizing an Unstable Economy* (1986) he argues further that it is *a characteristic* of the type of economy we live in in good times.

to the “production” of financial instruments. The extra liquidity, if used, will result in asset price increase, which again will lead to increased investment, thereby creating a positive feedback loop. These experiments with the extension of the debt structure can stretch over a number of years, during which the limits of the market are gradually tested. The revaluation, when something goes wrong, can be (and as a rule is) quite sudden and profound.

Minsky argues that, somewhat paradoxically, financial innovation and evolution are stimulated by attempts by central banks to curb inflation: new products and practices come in to supply extra liquidity. In the extreme, they might increase liquidity by more than what central banks have managed to reduce it. It is this excess liquidity — the liquidity above what the monetary authority deems appropriate — which is used to start a speculative fever.

Galbraith’s view on financial innovation is quite similar to the one of Minsky, but his wording is rather brusque¹⁸:

The rule is that financial innovations do not lend themselves to innovation. What is recurrently so described and celebrated is, without exception, a small variation on an established design, one that owes its distinctive character to the aforementioned brevity of the financial memory. *The world of finance hails the invention of the wheel over and over again, often in a slightly more unstable version. All financial innovation involves, in one form or another, the creation of debt secured in greater or lesser adequacy by real assets* [emphasis added]. (Galbraith 1993:19)

3.10. Speculation

Two instances of speculation have already been mentioned. Galbraith surveys speculation and provides some interesting features, which were present in the episode surrounding the recent crisis, even if that fact has not received much attention.

Mainstream theory views financial assets (widely understood) as proxies for consumption. Galbraith, on the other hand, forcefully argues that in reality items in asset markets are purchased for their potential to change price. In contrast, items in the goods markets are purchased mainly for consumption. Combined with the observed fact that financial assets can behave as Giffen goods, all this explains why there is a fundamental difference in the way market participants react to price changes for the two categories. Further, financial excesses are characterized by positive feedback loops: “[t]he speculation building on itself provides its own momentum.” (Galbraith 1993:viii);:

Speculation occurs when people buy assets, always with the support of some rationalizing doctrine, because they expect their prices to rise. That expectation and the resulting action then serve to confirm expectation. Presently the reality is not what the asset in question...will earn in the future. Rather, it is only that enough people are expecting the speculative object to advance in price to make it advance in price and thus attract yet more people to yet further fulfil expectations of yet further increases. (Galbraith 1995:108)

Another recurring feature Galbraith decries is the trend to mostly ignore speculation itself in the proposed reforms and regulatory changes – in effect, one treats the symptoms and does not try to cure the disease.

In view of these features, his observation that speculative episodes always stun experts and market participants is nothing but a necessary consequence of the reliance on a misspecified paradigm.

¹⁸ His works give the impression that he did not suffer from excessive sympathy for the financial sector and its practices.

3.11. Brevity of financial memory

This is more or less the essence of the “this-time-is-different” syndrome of Reinhart and Rogoff (2009). There are different views on the phenomenon. Galbraith argues that financial memory is extremely short: “There can be few fields of human endeavour in which history counts for so little as in the world of finance.” (Galbraith 1993:13) He believes that the maximum it can last is 20 years. Minsky, on the other hand, is of the opinion that financial memory does not fade so fast. Tylecote argues that the loss of financial memory is a result of the change of generations.

3.12. Psychological factors

There are various factors, which for different reasons cannot be adequately incorporated in economic models, but which, nevertheless, matter in the real world. While we may not be able to represent such factors analytically, it is, in my opinion, essential that we are aware of them and that they impose limitations to the applicability and usability to the real world of the unqualified results provided by economic theory, on which policy design relies. Unfortunately, this has hardly been the case for the past three decades.

One such example is psychological factors. All of the theories reviewed in this section, save the Austrian School's, concur in their importance, especially on the aggregate level. What is more, these patterns of behaviour appear to repeat themselves. Fisher identifies four phases in the psychology of going into debt¹⁹. They correspond more or less to the four stages identified in the Long Wave Theory. The latter, however, explicitly incorporates speculative euphoria as a process which develops during the third stage of a wave, known as a *plateau*: “In such times people are subject to a peculiar element of group dynamics to suspend caution because their peers have done so.” (Mager 1987:196) Galbraith argues further that euphoria is willingly sustained by the ones involved, usually as justification of the circumstances which make them rich²⁰. Thus, euphoria is driven by self-fulfilling prophecies, which act as a positive feedback loop. Minsky, Galbraith and the proponents of The Long Wave Theory argue that it is a necessary prelude to a financial crisis, and that it almost inevitably arises as the result of a successfully functioning economy. Lasting tranquillity over extended periods seems to affect the formation of expectations: the attitudes towards debt, risk-taking, etc. are subject to inertia and there seems to be strong bias to extrapolate the recent past. Combined with the “this-time-is-different” syndrome, this results in an increase in the proportion of the value and type of financial assets, which are sensitive to sharp revaluations of expectations. More specifically, with the fear of economic troubles weakened, risky behaviour on borrowed money increases. Gradually this increases the layers of private debt and thereby the latter's burden on the financial system, making it more susceptible to disturbances²¹.

These two traits of behaviour — subjecting oneself to euphoric group dynamics and falling prey to the self-serving bias — show that one of the major assumptions behind modern mainstream economic theory — that of rational expectations — is not substantiated empirically. Backed up by results obtained in the fields of behavioural and experimental economics, this provides further support for the necessity to bring economic theory closer to reality: according to Minsky the important change in this instance is the way choices under uncertainty are modelled. He believes that when economists construct preferences, they should use results from social sciences dealing with the psychology of uncertainty and the social psychology of waves of optimism and pessimism.

On the other hand, there may be good reasons for not including in economic paradigms certain factors which matter in the real world. For example, the aforementioned psychological factors, if not

¹⁹ See 47. in Fisher (1933:349) for details.

²⁰ This is an instance of what is known as self-serving bias.

²¹ To Minsky this was a reason for the turbulent times of his. With the benefit of hindsight, it can be safely claimed that the same process was at work in the run-up to the recent crisis.

in their entirety, often include a significant amount of non-rationality, which of course cannot be modelled when one of the building blocks of economic theory is the assumption that agents are rational. Maybe some insights from the field of behavioural economics will prove useful in overcoming this obstacle. Another example is fraud: while one can, in principle, model it as a form of rent-seeking, obtaining a good match of its real-world realization may be an impossible task.

3.13. Role of monetary policy

On the whole, the pre-crisis literature seems to be unanimous on monetary policy's being a source of trouble for the financial markets. The views fall into two categories. On the one hand, the representatives of the Austrian School and the adherents of The Long Wave Theory argue that (loose) monetary policy is the source of financial excesses²² for the former and the generator of a long wave through the credit cycle²³. On the other hand, Keynes and Minsky consider (loose) monetary policy only a necessary condition for the development of a financial boom, and Cooper argues that central banks can amplify boom-bust cycles over time. Whichever view may be the correct one, the fact remains that central banks' tampering with liquidity has had a negative effect on financial stability. It remains to be seen whether the recent crisis has provided new insights to monetary policy theoreticians and practitioners, so that the monetary authorities cease being a source of potential danger to financial stability.

4. The Recent Crisis

The recent crisis is such a vast topic that only some of its aspects — two of them to be specific — are relevant to the relatively narrow scope of the current composition. The first one is that in view of the material presented in sections 2. and 3., there is no doubt that the recent crisis was an endogenous product of the way the world economy has developed since the 1970s²⁴. The second aspect of the recent crisis, which is the more important for the purpose of the current composition and therefore is more thoroughly presented, is its causes. The importance stems from two facts. First, the identification of its causes substantiates the claim that the recent crisis was indeed an endogenously generated phenomenon. Second, knowledge of the roots of the problem must serve as a starting point for the development of both ex-ante and (outside the scope of the current composition) ex-post measures against future financial crises.

The material in this section, particularly heavily in subsection 4.4., builds primarily on Brunnermeier (2009) and Bean (2009).

4.1. Origins of the recent crisis

The magnitude of the recent crisis and its wider and long-lasting impact suggest that it was the culmination of decade-long development. It can be conjectured that the first seeds were sown with the onset of globalization and the rise in the importance of the financial sector as one of its major pillars in the 1970s. Imbalances had started to build up already in the aftermath of the collapse of the

²² "...that many things were done under the influence of artificial stimuli that would not have been undertaken without political and banking encouragement." (Schumpeter, *Business Cycles: A theoretical, historical and statistical analysis of the Capitalist process*, 1939, p. 265, quoted in Mager (1987:202)

²³ According to Tylecote (1992), this has been the case since around the turn of the previous century, when the monetary policy feedback mechanism started having a significant procyclical effect.

²⁴ It is beyond the scope of the current composition to discuss whether this outcome was the sole, unavoidable conclusion given the "concrete historical conditions" which characterized the time path of the world economy during the period. Still, with the benefit of our vantage point in 2014, there are strong indications that this seems to be the case.

Bretton Woods System, but together with increased reliance on debt²⁵ (both public and private²⁶) they did not come to the fore until the 1980s. At the same time the conditions for the thriving of financial engineering were constantly improving.

While the crisis surprised most of the mainstream (macro)economists, non-mainstream authors had both predicted that it will occur and that it will be the necessary and inescapable consequence of the aforementioned development. Consider, for instance, the following two quotes from Mager (1987) (mark the year!):

The world [emphasis added] is supremely sensitive to financial signals, as if it is waiting for a major financial crisis [emphasis added]. Such a crisis was narrowly averted in 1974, again in 1978 and 1980, and in 1983 and 1984. But each year, the debts continue to escalate.... (Mager 1987:238)

However, with the added straws of increasing debits by consumers, business, the federal government, and foreign debtors — the imbalances created in a free, increasingly deregulated market that make for an extremely fragile situation — we are apparently waiting for the dramatic crisis that will start the banking and speculative dominoes falling²⁷. (Mager 1987:239)

Another example is Minsky (1986), in which he opines that the reforms conducted during the Reagan years have worsened the prospects for financial stability, but that it would take time for their full effect to be felt. His prediction was that the foundations for another round of inflation crises and serious recessions were laid — the years to come proved him right.

4.2. The recent crisis fits into the historical pattern

It has been acknowledged in the post-crisis literature that the recent crisis fits well into the historical pattern: it was preceded by a build-up of significant imbalances²⁸: deterioration of current accounts and strong growth of debt²⁹, increasing consumption (often leveraged). The extensive use of leverage led to low capital adequacy, especially of the international banking sector. In addition, the more or less continual deregulation and financial engineering stood behind the prolific development and extensive use of complex financial instruments, often laden with hidden risks, which increased liquidity to an excessive amount. This led to underestimation of risk and powerful asset price inflation. The complexity of the financial sector increased, which made the localization of risk difficult. These developments made the financial sector as whole highly susceptible to reversals in asset prices and/or adverse changes in the ability to shoulder debt.

The aftermath of the crisis is also a collection of old acquaintances: precipitous decline in asset prices, contraction of markets for assets, waves of illiquidity and insolvencies and a significant and lasting impact on the wider economy in the form of depressed activity over an extensive period.

²⁵ Tylecote (1992) explains the observed fact that the incentives to borrow were not significantly affected by the high interest rates throughout most of the 1980s with the liberalization of the credit market and a “wealth effect” provided by a boom in housing prices (mostly in the UK, USA and Japan).

²⁶ Examples are provided by the deficit build-up during the Reagan years and the use of leveraged buyouts.

²⁷ This grim prediction is the final paragraph of his book.

²⁸ While the existence of imbalances has been acknowledged already before the crisis, no unanimity was achieved on whether they were problematic. The sovereign debt crisis in the Eurozone has brought the discussion on the topic to the fore, though.

²⁹ Already Tylecote (1992) identifies existing debt structures, terms of trade and inequality as the major imbalances in today’s global(ized) economy.

4.3. The recent crisis – a collection of crises

According to NOU (2011:1), there is agreement in the post-crisis literature that there was no single immediate cause (not to be mistaken with “trigger” here) for the recent crisis. Instead, there were many, each of which was quite serious in itself, but still none was strong enough to set off the crisis alone. For this reason, the recent crisis is regarded as a collection of country-specific crises. They combined with an array of imbalances and systemic weaknesses through positive feedback effects to transform the original problem in a segment of the US home market into the most serious financial crisis for decades. It was triggered by the collapse of Lehman Brothers³⁰.

4.4. Causes of the recent crisis

The multitude of causes notwithstanding, there seems to be agreement in the literature that excessive liquidity in the markets and improperly structured incentives are the main contributors to the recent crisis. There are at least three sources whence they came: current account imbalances, misplaced economic policy, of which inadequate regulation together with lax and deficient supervision have received most of the attention and blame for the recent crisis, and changes in the way financial institutions operate and how they earn their profits.

4.4.1. Current account imbalances. Current account imbalances are the result of the gradual build-up of trade imbalances which the changes in the global economy since the early 1970s have contributed to. These became significant after the year 2000³¹. The necessity to finance current account deficits was met with the money which had been accrued as a result of the high savings rates in the growing economies: huge inflows of money found their way into the USA in search of profitable returns there, increasing liquidity and pressing interest rates down.

4.4.2. Policy failure. In the years before the crisis it appeared that economic policy had finally moved away from being guided by emotions to being steered by what appeared to be solid science. At the same time, the odious politicians had finally renounced their right to determine the goals and the course of economic policy and had, instead, become just executors of what sages recommended. Yet, the recent crisis shattered this idyllic picture. The authors who have delved in the roots of this disappointment concur that flawed economic policy took three broadly discernible forms: regulation and supervision proved inadequate, too much was relied on economic models mistakenly believed to be virtually exact replicas of the real-world economy, there was no consensus on how the extended period of low interest rates should be regarded and handled.

As far as the first form is concerned, the financial crisis revealed severe deficiencies in the regulation of both financial institutions and markets internationally. Inadequate regulation provided ample opportunities for the financial firms to make use of what has been termed “regulatory and ratings arbitrage”. The financial firms, having the “brightest people” (outside central banks) on their payrolls, took full advantage and reaped whatever they were able to. This led to the significant reduction of the financial system’s capital adequacy on the eve of the crisis. As a result, quite a few examples of regulatory failure have earned their place in history books: banks were able to move funds off their balance sheets to various vehicles, while still meeting capital requirements, at no loss of their credit ratings; regulatory loopholes allowed the institutions of the “shadow banking system”³² to engage in

³⁰ Note that this accords nicely with The Long Wave Theory’s description of the panic, which ends the plateau (third) period of a wave: the precursor of each panic is monetary phenomenon which, *seen in isolation*, is of minor importance. Its impact is due to the reversal in public confidence it causes.

³¹ As a matter of fact, trade imbalances were accorded significant attention internationally, and singled out as a major source of risk in all but a few macroeconomic forecasts made during the first decade of the present century. However, the significant attention was balanced with equally significant inaction to address the problem prior to the crisis.

³² The term “shadow bank” is used to denote any financial institution which out of its operations is de facto a bank, but out of its statute is not so de jure.

the same form of maturity transformation — financing long-term assets by short-term debt instruments — as ordinary banks. However, because they did not count as banks on paper, they were not subject to the same strict regulation as ordinary banks, which allowed shadow banks to extend more loans.

The ones entrusted with the supervision of the financial markets failed thoroughly. It is the failure to oversee the systemic risk in the financial system as whole that will remain the lasting symbol of their poor performance. To their defence, it must be admitted that they were organized in a way such that they both prior to and under the crisis, were neither able to prevent nor to deal with a crisis of such proportions.

Regarding the second form of policy failure — model inadequacies — Keen (2012) argues that key intellectual issues in the financial sector have been ignored for too long. It will be no exaggeration to argue that all the deficiencies of economic models exposed by Minsky, Galbraith and Cooper received empirical confirmation in the recent crisis.

So, what did it demonstrate to us? First, efficient, well-working markets cannot be taken for granted. Second, the way credit is modelled in macroeconomics need a major change: the assumption of complete financial markets has to be abandoned in favour of accepting the importance frictions in the financial system play in general. Third, the crisis exposed a feature of financial intermediaries inconsistent with theory: instead of the black-box-approach, which regards them as some simple entities devoid of incentive and information problems, they have to be seen as an additional layer and a major potential source of economic trouble. Fourth, the importance of high debt and imbalances must receive more attention. Fifth, calculation of risk-weighted capital requirements must be improved, so that it is ensured that the risk-bearing ability reflects the true risk in a given institution's activity³³.

What about the third form of policy failure? Well, there seems to be agreement now, *after* the crisis, that interest rates were too low for too long. This fact may be explained by the effects of *The Great Moderation* and, to a lesser extent, by the workings of monetary policy. *The Great Moderation* is a term used to describe the period of twenty or so years from the mid-to-late 1980s until the recent crisis, which was characterized by low volatility over the business cycle in the advanced economies. Gradually it came to be perceived by the market participants as a permanent change, which, in turn, started pushing risk premia down. The important (and natural) thing was the way economic agents reacted to this development: their willingness to take on risk increased³⁴. Then the turn of the century saw the dot-com bubble burst and the conventional loosening of monetary policy to help counteract the negative consequences. Low policy rates by central banks are not by themselves a major reason for debt build-up — echoing Keynes, funds must not only be available, but spent as well. Overall, central banks accorded far too little attention to the increase in credit and asset prices in their policy decisions. Two explanations for this are put forward: first, central banks did not have explicit mandate to prioritize financial stability; second, it is claimed that, in practice, it is often impossible to distinguish between situations where asset price inflation is due to rational adjustment to new information about fundamentals, and situations where asset inflation is the result of a bubble — this is often cited as one of the main justifications for inaction by authorities in the face of such

³³ One necessary improvement is that the models employed should make use of data also from troubled periods. The ones used before the crisis relied on historical data from a relatively tranquil spell. The historical correlations, which formed the fundament for the assumed extensive risk diversification, failed to hold once trouble in the markets arose. Another necessary improvement is increased adaptability of models to new financial instruments — a weakness of many pre-crisis models was a lack thereof.

³⁴ Note that this fits nicely with Minsky's theoretical prediction and later empirical observation that a trend in the post-war period has been that the proportion of speculative and Ponzi financing tends to increase as the period without a serious depression is extended. Note also the effect of lengthy tranquillity on expectations.

developments. The first explanation reflects the belief in the efficiency of financial markets and their ability to self-regulate. The second one is a manifestation of the limitation on our knowledge of the way the economy works.

4.4.3. Changes in the financial sector. The process of financialization which has characterized the development in the financial sector has led to profound changes in the latter. Following the practice in economics of searching for analogies with phenomena in other fields, I dare draw a parallel with biology: the changes of the financial sector were both morphological — affecting its structure, organization and size — and physiological — involving shifts in profit structure, operation and financing.

As far as the first set of the aforementioned transformations is concerned, the recent crisis reminded us of one of the downsides of the mutual interdependence in a world of growing economic activity and increasingly intertwined financial markets: a localized crisis can spread relatively quickly through a multitude of channels and affect adversely everybody.

The interdependence among the participants in the financial system had been steadily increasing in the years before the crisis. This created a fragile giant net of cross-holdings of underlying loans, so that much of the risk in the issued assets remained within the financial sector. The vulnerability of this structure stemmed from the positively correlated returns on the loans and the systemic nature of the main risks. With institutions' being highly exposed to each other, a lot of counterparty risk lay hidden. The risk exposure of others was not a concern while the bubble was inflating, but once the solvency of a number of institutions came into doubt (regardless of whether it was for fundamental or speculative reasons), uncertainty about others' positions caused trust between financial institutions to all but evaporate. As a result, liquidity dried out.

Another result of market integration was the steadily increasing concentration in the financial sector. In this changed market structure many financial institutions increased in size and complexity. A negative side of this increased concentration, which was highlighted by the recent crisis and which is in need of being addressed, is that it makes a larger portion of the financial sector systemic. The recent episode of trouble only served to exacerbate the problem: the difficulties in many financial institutions meant that their most realistic way to survive was by being merged or bought out, thus allowing new surviving institutions to become even bigger.

The preceding two paragraphs serve as conclusive empirical confirmation of Minsky's conjecture that increased layering of obligations increases the fragility of the financial system.

The second set of transformations — changes in profit structure, operation and financing — did not go unnoticed (but did go unaddressed): already Allen and Santomero (1999) highlight the shift in banks' source of income from an interest-rate-spread to a fee-producing one. This directly affected their incentives towards fee maximization, which did not necessarily go together with the optimal choice of the debt-liability mismatch. This development reached its pinnacle with securitization. Securitization can be defined simply as the decoupling of loan origination from loan monitoring. In theory, it was beneficial to all. As far as financial institutions are concerned, the banks originating the loan could profit from at least two sources: first, they did not have to keep the loan (and hence the associated risks: default, interest rate, counterparty) for its entire maturity, which alleviated their debt-maturity mismatch; second they could keep the proceeds from selling the loan at face value. Both of these gave the banks incentives not to care (very much, if at all) how (un)likely the potential borrowers were to repay the loan, only there be as many of them as possible, thus making the banks extremely willing to extend loans.

There has been a tendency for horizons in economic life to become shorter³⁵. Nowhere is this more obvious than in the setup of remuneration schemes. Even though payment arrangements favouring short-termism are present in many sectors, the ones in the financial sector are a class of their own. Even prior to the crisis it was obvious to many that the remuneration schemes of key players appeared to reward short-term returns, while no heed was taken of potential losses over the medium and long term. The ones concerned had every incentive to take on too much risk: if the good state of nature prevailed, they would appropriate (and did appropriate) almost all of the gains, while they would incur only insignificant losses when things turned sour (which according to the ruling theory was very unlikely). Further, true to the principle of success's having many fathers and failure's being an orphan, they could always claim (and did claim) that all success was the result of their actions and hence it was just to accord to them the lion's share of the gains; and, of course, could claim (and did claim) that failure was due not only not to their actions, but even in spite of them, which again entitled them to sky-high remuneration (and this was bought).

It will be no exaggeration to say that the single most important factor behind the spread of the recent crisis was the way financial institutions financed themselves³⁶. Virtually all participants in the financial sector were made painfully aware of the dangers posed by excessive leverage³⁷ and reliance on short-term financing. The low price of risk in loan markets and the favourable treatment of debt for tax purposes skewed many institutions towards overleveraging. The use of short-term financing had been steadily rising in the years before the crisis³⁸. Reliance on it posed significant risks, because it made financial institutions extremely susceptible to sudden squeezes in liquidity. This was not a problem as long as the markets were full to the brim with liquidity, or as the then-chief of Citigroup Chuck Prince was quoted by *The Financial Times* to have put it in early July 2007 "When the music stops, in terms of liquidity, things will be complicated. But as long as the music is playing, you've got to get up and dance. We're still dancing." (*Financial Times* 2007). So one day the music simply stopped. Thus, the recent crisis has proved Minsky right.

4.5. The recent crisis in Norway³⁹

The recent crisis, similarly to previous ones, did not affect all countries equally; for instance Norway enjoyed a less bumpy ride during and after it, which, according to NOU (2011:1) can be attributed to luck, competence and caution.

There are three aspects to luck identified in the report. First the size of traditional industry as a fraction of GDP is small⁴⁰ and a significant share of its production is supplied to the petroleum sector. Second, there was no significant and lasting decline in the price of oil, so demand from the petroleum sector remained stable. Third, the high debt of households and the wide-spread use of

³⁵ The trend was noted already by Minsky, who explains it with the more and more strained liability structures which resulted from the steadily increasing reliance on debt.

³⁶ Already Minsky in "Finance and Profits: The Changing Nature of American Business Cycles" (Minsky 1982:14-58) surveys the changing nature of the financing of US banks. See also Allen and Santomero (1999) for an update on the development.

³⁷ The risk of high debt was alarming already in the 1980s to a degree that *Business Week* had devoted its cover to label the economy "The Casino Society" in its 16th September 1985 issue, quoted in Mager (1987:230):

No, it's not Las Vegas or Atlantic City. It's the U.S. financial system. The volume of transactions has boomed *far beyond anything needed to support the economy* [emphasis added]. Borrowing — politely called leverage — is getting out of hand. And futures enable people to play the market without owning a share of stock. *The result: the system is tilting from investment to speculation.* [emphasis added]

³⁸ For example, Brunnermeier (2009) mentions that the amount of repo financing doubled between 2000 and 2007, corresponding to an average growth of 10-11% per year.

³⁹ See NOU (2011:1) for a full survey.

⁴⁰ The report gives the figure of 14% for the period 2003-2006.

flowing interest rates ensured faster and stronger effects of changes in interest rates on income and hence demand.

Competence can be attributed to the way the country's real economy is organized. More specifically, as NOU (2011:1) argues, a large public sector, relatively equal income distribution, high degree of unionization, relatively strongly centralized and coordinated wage setting and a comprehensive social safety net are among the features which ensure stable demand for goods and services, while the way the proceeds from the petroleum sector are managed provided the country with solid public finances, which allowed considerable room for manoeuvring when powerful stimuli were provided by the authorities.

Caution refers to the way the financial sector in Norway is organized and the framework within which it operates. NOU (2011:1) gives credit to stricter regulation: it precluded regulatory arbitrage, subjected non-banks performing tasks typical of banks to the same regulation, prevented financial institutions from using off-balance sheet tricks, thus ensuring that the capital adequacy of Norwegian financial firms at the beginning of the crisis was quite solid⁴¹, at least relative to their peers elsewhere. In addition, Norwegian banks' main source of finance is deposits, and they happened to hold few securities, of which only a tiny fraction suffered a precipitous fall in value.

5. Containing Systemic Risk – a Task of Macroprudential Policy

5.1. Macroprudential policy and macroprudential regulation

The recent crisis showed that the costs of the macroeconomic harm caused by the vagaries of the financial system are enormous. In addition, according to Laeven and Valencia (2010b) the bill, in terms of output losses and increases in public debt, was more massive on average compared to its counterparts from previous crises, even though the direct fiscal outlays were smaller⁴². This, according to NOU (2011:1), implies that there are significant gains to reap by reining the vagaries of the financial system in, benefiting all actors in the economy and public finances. The policy which tries to achieve this goal is known as macroprudential policy. The Bank of International Settlements describes macroprudential policy as “the use of prudential tools with the explicit objective of promoting the stability of the financial system as a whole, *not necessarily of the individual institutions within it* [emphasis added]. The objective of macroprudential policy is to reduce systemic risk by explicitly addressing the interlinkages between, and the common exposures of, all financial institutions, and the procyclicality of the financial system” (quoted in NOU (2011:1, p.202)). Further, it aims “to limit system-wide financial distress that stems from the correlated exposure of financial institutions and to avoid the resulting output losses in the economy” (Korinek 2011a:30). Regulation based on such an approach, concerned with the overall risk in the financial system (i.e. systemic risk), is called macroprudential regulation. In contrast, microprudential regulation is concerned with the stability of individual entities.

Acharya, Pedersen, Philippon and Richardson (2010) acknowledge that the pre-crisis regulation was skewed heavily toward microprudence, because stability at the micro-level was considered the key element – in other words, it was assumed that financial stability is just the sum of stable financial

⁴¹ Good returns in the years preceding the crisis further strengthened the soundness of financial institutions.

⁴² For instance, IMF (2010) provides the following figures for what they term “advanced G-20 economies”: average non-recovered costs of direct fiscal support as of end-2009 is 2.8 percent of GDP, cumulative output loss in countries which went through a systemic crisis is, until the report was published (September 2010), 26 percent of GDP, while government debt of these same “advanced G-20 economies” is projected to increase by 40 percent on average over the period 2008-2015

institutions. The financial crisis has shown, however, that there is not just a quantitative, but also a qualitative aspect⁴³ to financial stability, so that it is more than the mere sum of stable financial institutions. Jeanne and Korinek (2010) argue that the belief that microprudential measures are sufficient to ensure financial stability on the aggregate level has its roots in deficiencies in most models considered in the literature – they do not satisfy the conditions necessary for the existence and uniqueness of equilibrium, which, in addition, is derived by iterating the models' first-order conditions⁴⁴. Instead, they argue that there is a multiplicity of equilibria, especially relevant for highly leveraged institutions, attributable to a self-enforcing loop linking consumption to the price of collateral, which is the financial magnification of shocks and it can lead to self-fulfilling asset price crashes, if it is sufficiently strong.

5.2. Systemic risk, financial amplification and procyclicality

Acharya et al. (2010:2⁴⁵) define “systemic risk...as the failure of a significant part of the financial sector — one large institution or many smaller ones — leading to a reduction in credit availability that has the potential to adversely affect the real economy.” It is a realization of a state of the world in which the financial sector as whole becomes undercapitalized: financial constraints bind and financial amplification effects are triggered, impairing the process of financial intermediation as a consequence. In NOU (2011:1) one finds that the Bank of International Settlements distinguishes between systemic risk which is built up over time as asset prices and debt grow fast, and systemic risk which is due to the various forms of interdependence between the financial institutions. Systemic risk tends to be highly non-linear on the downside – a feature which makes it difficult to model it. Weder di Mauro (2010b) argues that the extent of propagation of the initial shock, identified as the failure or some other form of distress of a single institution⁴⁶, and its impact define systemic risk – in effect reiterating the ideas of Minsky and The Long Wave Theory, albeit in different wording.

What factors contribute to increased systemic risk? Korinek (2011a) lists four: higher initial financing requirement by bankers (because then it is optimal for them to incur some fire sales in low states of nature), lower endowment of households (it means reduced supply of capital to bankers), lower minimum return in strained periods which cannot be used as collateral, and greater relative risk aversion of households. Further, NOU (2011:1) notes that systemic risk increases with the debt level, while Acharya et al. (2010) point out the presence of a lot systemic risk in the over-the-counter markets, from which it can be concluded that higher proportion of trading being conducted there is the next factor. Lastly, Weder di Mauro (2010b) argues that the incentives of the public sector also matter, because they tend to suffer from time-inconsistency, which according to Weder di Mauro and Klüh (2010) has become worse over time. The latter study also provides some justifications for the authorities' forbearance which time-inconsistency engenders.

A model in Korinek (2011a) shows that market participants find it privately optimal to take excessive amount of systemic risk. As a result, the economy suffers from too much financial amplification: the process by which the systemic interaction between debt accumulation and asset prices contributes to magnify the effect of booms and busts, and which is thereby important in many economic settings⁴⁷. It is a reason for the procyclicality of financial markets. It also amplifies the initial

⁴³ One such aspect is the non-rational traits of behaviour, such as mood swings and “market sentiments”, discussed in subsection 3.12.

⁴⁴ Jeanne and Korinek (2010) argue in general that the models solved by linearization used in the literature are less suited to study systemic crises than regular business cycle fluctuations.

⁴⁵ The version of their paper I have used is not paginated, so the quote is from what is the second page, including the title page with the abstract of the article in the count.

⁴⁶ Thus corresponding to what Minsky and the adherents of The Long Wave Theory call *the trigger*.

⁴⁷ That notwithstanding, Jeanne and Korinek (2010) mention that there are few welfare analyses of optimal policies addressing the issue.

disturbance (*the trigger*). The latter spreads through two main channels of propagation, identified in Weder di Mauro (2010b) and Acharya et al. (2010): a direct one, which manifests itself in counterparty risk (i.e. the connections between institutions), and an indirect one, which manifests itself in fire sales (i.e. the spillover from forced asset sales).

Shin (2010) argues that the recent crisis has shown convincingly that banks are not passive institutions (“black boxes” as I referred to them in subsection 4.4.), but both serve as the engine of financial cycles through the active management of their balance sheets, and amplify the financial cycle through their lending decisions. Especially, leveraged institutions play an important role in the latter process: the high vulnerability to erosion of their capital makes them very sensitive, so that even small shocks can cause them to engage in substantial adjustment of their asset portfolios.

Measures of systemic risk have been proposed. Shin (2010) has put forward the idea to use the stage of the financial cycle as an indicator of systemic risk, based on the positive relation between the two. He argues that the composition of liabilities of the banking sector reflects the former. He, therefore, proposes the use of the ratio of non-core to core⁴⁸ liabilities as a useful indicator of the stage of the financial cycle and the implied degree of vulnerability of the banking system to adverse developments in the financial sector. His proposal is based on the fact that the core liabilities of banks can only grow at the same pace as the growth of aggregate wealth in the household sector. In a boom the increased demand for credit cannot be supplied by enough deposits, so banks have to resort to other sources. Hence, the larger the proportion of non-core liabilities⁴⁹, the greater the boom and therefore the greater the systemic risk. Shin (2010) argues that the composition of liabilities is a better early warning signal of vulnerability than conventional asset side indicators, such as non-performing loans or capital ratios. An additional advantage is that this indicator will show the stage of the cycle also on the aggregate level. Based on US data, Shin (2010) argues that if the ratio of non-core (in his example repos) to core liabilities (in his example M2) had been tracked more closely, its rapid increase before the crisis could have alarmed decision makers that swift, resolute countermeasures were necessary.

5.3. Challenges to macroprudential policy

It is important to be aware of the limitations of macroprudential policy and to have a sober view of what it realistically can achieve. Echoing the ideas of Minsky and Cooper, Kocherlakota (2010) states that no regulatory structure is able to completely eliminate the factors which can bring about a crisis, while Weder di Mauro (2010b) also argues that even after measures to deal with systemic risk are implemented, there will still remain some residual risk of a systemic crisis. Thus, it is natural that one of the lessons of the recent crisis is that overambitious policy goals should be abandoned, together with the overvaluation of the abilities of authorities to fine-tune activist stabilization policy⁵⁰.

Lack of knowledge of the exact workings of the economy and measurement problems complicate the timely implementation of necessary measures. According to NOU (2011:1), the higher the policy

⁴⁸ Shin (2010) defines core liabilities as the claims of ultimate creditors (households) on the financial intermediation sector. Archetypical examples of such claims are retail deposits and money market funds. The author points out that not all deposits should qualify as core liabilities, however. If a firm issues some securities and then deposits the proceeds into a time deposit account, the latter should not count as a core liability to the bank entrusted with it, because the depositor (the firm) is not the ultimate claim holder. On the other hand, core liabilities can include items which are not technically deposits: e.g. covered bonds issued by banks or some types certificates of deposit. As far as non-core liabilities are concerned, they are exemplified by wholesale funding.

⁴⁹ The growth of non-core liabilities is often accompanied by shortening of the maturity of liabilities, which, if taken into account, apparently can improve the quality of the indicator.

⁵⁰ For example, Beans (2009) concedes that there has been more faith by both the public and policy makers in the omnipotence of the latter than what reality provides for.

ambitions, the surer one must be of the condition of the economy and its development further ahead. The recent crisis has been but the latest reminder that our ability to (gu)e(s)stimate, where the economy is at a particular point in time and where it is heading for, is limited.

Knowing that the effects of systemic risk materialize through financial amplification effects would suggest that the potential first-best policy would be to break the amplification effect, because the real world financial markets are characterized by frictions (due to imperfect and asymmetric information) which cannot be eliminated completely and hence the operation of the financial accelerator cannot be prevented. This breaking is implemented either through injection of liquidity into constrained firms (popularly known as “bailouts”), or through stabilizing asset prices (“protecting” in Minsky’s terminology) by buying up fire sales (popularly known as “socializing the losses”). Both measures, however, create huge moral hazard concerns, which often end up contributing to a crisis, as pointed out in the literature.

Korinek (2011a) notes that there are some “perverse” economic implications from the operation of financial feedback loops; for instance, the observation that the demand for assets is an increasing function of their price, which as noted before is the description of a Giffen good. Such issues command caution when contemplating what measures to apply.

6. General Considerations – Taxes and Regulatory Command as Tools for Implementing Macroprudential Policy

First of all, a clarification of the terms used is necessary. Strictly speaking taxation and regulatory command are just two subtypes of regulation. Nevertheless, in the following pages, unless explicitly otherwise specified, I will use interchangeably “regulation” and “regulatory” as a catch-all term to refer to any measures other than taxation.

The financial crisis spurred a discussion about the use of taxes and fees as means of addressing the weaknesses in the financial system it exposed, as well as of diverting income away from it. Still, according to Keen (2012), the ongoing process of financial sector reform is dominated by a full regulatory agenda, reflecting the dominant consensus in the literature that changes in regulation are the way to go, because it is believed to be the most effective measure and has, in addition, the benefit of a long tradition, both internationally and in each country. Keen and de Mooij (2012) caution, however, that this does not imply that regulation is the socially optimal way to implement macroprudential policy. The consideration of other measures, such as taxation, therefore, is important. The reason is the practical infeasibility of the theoretical result which says that, under the too restrictive conditions of full information and no concern for revenue, the two approaches are equivalent and relying solely on either one is enough to achieve the social optimum. This means that “[in t]he real world...the choice of instruments is...a substantive one” (IMF 2010:10).

6.1. Differences between taxation and regulation

Three differences between taxation and regulation are closely related to each other. Taxes can affect the risk in the financial system by changing the incentives faced by the economic actors, i.e. through the substitution effect. Regulation, on the other hand, changes the opportunity set they face. More generally, corrective taxes can be employed against specific, identified sources of risk and other concerns in a given financial institution. Closely related is the use of corrective taxes as automatic stabilizers of the financial system and as a means of mitigating its procyclicality – they can thus be incorporated in the general use of taxation as a form of automatic stabilizers over the business cycle. Thus they can contribute to increased flexibility and provide greater room for manoeuvring,

especially on the national level, which, according to NOU (2011:1) is important for the now cliché small-and-open economy, particularly in view of the fact that global coordination of measures has so far proved an insurmountable hindrance. That notwithstanding, IMF (2010) argues that there is greater scope for effective international coordination on regulation than on taxation, by virtue of existing institutional arrangements.

According to Shaviro, Shackelford and Slemrod (2011) and NOU (2011:1), taxation is better suited than regulation to meet the goals of macroprudential policy in the presence informational advantage enjoyed by the regulated over the regulating. On the other hand, IMF (2010) argues that in the instance of reducing risk-taking, regulation may be the better choice, as it need not have the information necessary for the imposition of a sufficiently⁵¹ progressive tax.

Keen (2012) and Weder di Mauro (2010b) point out that while both taxation and regulation can achieve the reduction of systemic risk, the former transfers funds to buffers under the control of the public sector, while the latter leaves them in the hands of the private sector. By reducing the net surplus of financial institutions by more than does regulation for a given capital ratio, taxation weakens private, but strengthens public buffers. Where the buffers should be located depends on the nature of the shocks which they are intended to cushion against: for instance, with negatively or weakly positively correlated shocks public buffers acts as pools of risk providing the cheapest way of dealing with institutional failures. IMF (2010) cautions, however, that with deficient resolution mechanisms public buffers are a source of moral hazard. On the other hand, private buffers contribute to reduced probability of distress or failure, which is especially beneficial as far as challenging-to-resolve institutions are concerned.

Another difference exists, according to the aforementioned reference, in the advantages each of the two approaches confers in an environment of uncertainty: taxes provide greater scope for response by the private sector, while regulation produces surer outcomes.

An important advantage of taxation is that it reduces the scope for discretion and with it the risk of regulatory capture, as IMF (2010) argues.

The Mirrlees Review (2011) points out one more advantage of corrective taxes: the tax revenue can be spent to reduce the tax burden in (an)other market(s), thus alleviating distortions in at least two separate markets.

6.2. Common features to taxation and regulation

Most of the features both approaches to reduce systemic risk share are the limitations they face in practice. First, any paper dealing with the issue stresses concerns about competition as the major factor which constrains the freedom of policy action. Second, uncertainty poses a challenge to both taxation and regulation, because policy must be set before the environment of operation is known. Combined with high model dependence, which makes it very hard to come with some general solution, applicable to as many situations as possible, it is likely to lead to a systematic preference for either method. Third, imperfections in design and implementation⁵² open up the possibility for avoidance. According to Weder di Mauro (2010b), quantitative regulation is bound to be ineffective, because more often than not regulatory changes are spurred by some accident, such that they are normally reactive and should protect from reoccurrence of known problems. Further, the ingenuity

⁵¹ Insufficiently progressive taxes will, according to IMF (2010) strengthen the incentives to take risk, by “leaving equity holders, and potentially key employees, with less ‘skin in the game’” (IMF 2010:53).

⁵² NOU (2011:1) provides as an example of poorly designed regulation the capital coverage rules in the Basel II framework: concerns have been voiced that the requirements are easier to meet in an upswing and can be quite demanding in a downturn, such that the tendency of banks to behave procyclically is actually strengthened.

of the financial sector has managed to find ways to exploit the less regulated parts of the system previously. There is little doubt that corrective taxation will not be immune to such deficiencies. Fourth, whatever measures are implemented, it is essential that they affect the behaviour of the intended target group – for instance, as far as taxation is concerned, Shaviro et al. (2011) point out the importance of the theory of incidence for ascertaining that the residual payers of the tax(es) are indeed the intended ones.

6.3. Role for corrective taxes and fees

The different properties of taxes and fees compared to regulation imply that in certain cases they can be a good supplement or even an alternative to it as instruments of ensuring that the financial sector is robust and serves its purpose in the wider economy, and not (just) its own narrow interests. Taking this into account, it is the legal and practical feasibility, and what is deemed or appears appropriate in a particular situation which will determine which of the two methods or a mix of both is the preferred tool to be used. Such a view is expressed, for instance, in NOU (2011:1), Shaviro et al. (2011)⁵³, *The Mirrlees Review* (2011)⁵⁴ and Cottarelli (2010), while IMF (2010:3) argues that coordination of taxes and changes in regulatory command are “critical for ensuring policy coherence”, yet sees the role of taxes as supplementing regulation. The optimal balance between taxation and regulation is a fruitful field of study: Shaviro et al. (2011) decry the lack of appropriate and adequate discussion of the optimal coordination of the two methods in the extant literature.

It goes without saying that if taxes and fees (potentially new ones) are to play any role, they must address the issues revealed by the recent crisis. More specifically, within the context of the current composition, they must contribute to altering the incentives faced by the actors in the financial system in a socially Pareto-superior direction, a consequence of which will be the likely change in the areas of activity of financial institutions and their role in the economy.

7. History of Proposed Corrective Taxes before the Recent Crisis

The idea of using taxation as a means against the excesses of the financial sector is not new. As mentioned in subsection 1.2. already Keynes, having had the opportunity to experience personally the whims and oddities of the way financial markets (mal)functioned and the tremendous amount of damage they could inflict over most of the world at the time, proposed in his *General Theory* (1936) the introduction of a tax on financial transactions to curb the speculative activities of Wall Street, which, he pointed out, were the primary motives for investment across the Atlantic⁵⁵. He also

⁵³ Shaviro et al. (2011) rationalize the continued use of regulation:

For two distinct reasons, therefore, the tool of regulatory command is likely to continue being necessary, in addition to any tax instruments that are adopted. First, expected social harm, other than the purely pecuniary to the government as insurer, is multi-dimensional and difficult to measure. Second, even that pecuniary harm cannot be measured entirely accurately through a risk adjusted-fee. Thus, the classic tax-or-regulation debate is surely beside the point with respect to financial institutions, because regulation of the financial sector both is not going away, and should not. In particular, capital adequacy regulations...will surely continue to be necessary, and indeed must presumably be improved and made harder to avoid given their failure in many cases during the period leading up to the financial crisis. Given, however, the multi-dimensional character of systemic risk, multiple tax and/or regulatory instruments are likely to be needed. (Shaviro et al. 2011:14)

⁵⁴ The authors behind the report argue that there are cases where direct regulation (for example, through bans, impositions) can achieve a better outcome, because pecuniary incentives need not be enough; this opinion is also shared by NOU (2011:1). Further, they point out that taxation may actually punish the prudent and fully-rational individuals

⁵⁵ One might consider him fortunate not to have lived to see this become the global standard of investing.

pointed out the difficult trade-off financial markets imposed on individuals and governments: the price for the abundance of liquidity is the heightened prospects of speculation.

The first concrete proposal for a corrective tax was put forward by James Tobin — and hence became known as “Tobin tax” — as a reaction to the turmoil following the collapse of the Bretton Woods system, which allowed speculation in the foreign exchange markets to thrive. By altering the trade-offs potential speculators face, it is supposed to reduce the rents accruing to them in the foreign exchange markets, thereby giving them incentives to “substitute away” from speculation. By its design it is an indirect, flat, ad valorem, (financial and more specifically, currency) transaction tax, which would be levied on all spot movements between any two currencies and ideally would be internationally agreed upon and uniform across the globe. Tobin argued that back in those days disturbances in the international financial markets were transmitted via currency exchanges. The significant costs to governments and national economies of offsetting these huge movements⁵⁶ — “financial transactions disguised as trade” in Tobin’s own words (Tobin 1978:159) — justified, in his view, “throwing sand in the wheels” (Tobin 1978:154) of the international money markets, even though he regarded it as an inferior option. It involves a trade-off, where some efficiency has to be sacrificed for some stability, so that larger inefficiencies to the real economy from turmoil in the financial markets are avoided. Finding the optimal balance requires that tax rate be sufficiently high to discourage speculation and, at the same time, sufficiently low lest it hamper trade in goods and services and productive long-term investment, through its effect on liquidity. Spahn (1995) did not think that this optimal balance could be found: either the tax rate would be perceived as too high, impairing financial operations, or as too low, in which case its effect on speculation would be insignificant. He further dismissed the Tobin tax on the grounds that there was no sure way of distinguishing ordinary trading from speculative trading, so that the result of the stability-speculation trade-off is uncertain. Theoretical and empirical studies of the effect the tax has on volatility do not come to any unambiguous conclusion. Further, there does not seem to be any consensus to date on whether the introduction of the tax is technically feasible, neither on how to limit the scope of avoidance and the damage from arbitrage opportunities it can cause. All this seems to explain why in *The Mirrlees Review* (2011) it is argued that the present scope of the tax, either in its original form or in some variant of it, is primarily for revenue-raising.

8. Financial Sector Externalities

The recent crisis is a grossly inefficient allocation which obviously violates The First Main Theorem of Welfare Economics. From this realization it follows that one or more of its prerequisites were violated. Whenever that is the case, it is customary in economics to speak of market failure – and the failure of the financial markets was an epic one, indeed⁵⁷. One of the violated prerequisites is the presence of externalities – Shin (2010) and Acharya et al. (2010) are among the studies which have acknowledged that the recent crisis has demonstrated that the financial sector exerts significant

⁵⁶ This transition to “financialization” is quite well visible in the following excerpt from Asbjørn Rødseth’s *Open Economy Macroeconomics* (2000):

Until some years ago most theories of the foreign exchange market...tended to assume that a major part of the supply and demand for foreign currency arose from the needs of importers and exporters....Today, however, most trade in foreign currency is *capital movements*. Foreign currency is bought and sold in order to change the composition of asset portfolios. Large stocks of wealth can be shifted from one currency to another literally in seconds. (Rødseth 2000:10)

⁵⁷ Acharya et al. (2010) attribute the inefficient allocation from the crisis to regulatory failure as well. From the brief account on it, and in general on policy failure, it seems that regulatory failure did not simply contribute to the crisis, but also exacerbated its scope and impact. This lends support to Minsky’s warning that implementing policy actions based on inadequate knowledge of the way the economy behaves can actually make things worse.

negative externalities on the rest of the economy, while Korinek (2011a) and Jeanne and Korinek (2010) present formal treatment of the phenomenon⁵⁸. More specifically, regulation failed to force internalization of the relevant costs, because it did not manage to close the significant gap between private and social cost of the financial sector's activities. As a result, the financial sector became too big and the actors within took too much risk. This gap is nowhere to be seen better than in the way the crisis was resolved: owners and managers of financial institutions, who had benefited from the circumstances which brought about the crisis, had to stand for disproportionately small fraction of the resolution costs. According to IMF (2010), the externalities can be amplified by the inherently procyclical nature of the financial system and the presence of big and complex financial institutions.

8.1. Types of externalities

By its very nature, systemic risk is a negative externality. It is a collective term for the realizations of what IMF (2010) calls macroprudential⁵⁹ externalities, several of which have been identified in the literature.

Pecuniary externalities, discussed in Korinek (2011a), are the key ones. They are defined as the effect an agent's actions have on the welfare of other agents through prices — for instance, when the former engages in fire sales or in bidding up the price of some asset — and are present when economic agents are subject to financial frictions. The importance of pecuniary externalities stems from the fact that they can lead to the realization of the main inefficiencies revealed by the recent crisis: excessive leverage, excessive reliance on short-term debt, excessive risk-taking. Further, they reduce the efficiency of the distribution of capital.

An instance of pecuniary externalities, of which Korinek (2011a) and Kocherlakota (2010) speak, is the fire sale externality. It is the result of a given institution's failing to internalize the effect of aggregate price decline on the value of other institutions' assets and collateral, when it engages in massive sell-offs, especially in times of trouble. One of the results in the model of Korinek (2011a) is that even the presence of complete markets to insure against systemic risk is not sufficient for internalization of the benefit of limiting economy-wide fire sales — the agents do not find it privately optimal to insure up to the socially optimal level; another one is that for the same reason in the presence of this externality agents face socially suboptimal incentives to raise additional capital.

The second type, systemic externality, is discussed in Kocherlakota (2010), Keen (2012), and Jeanne and Korinek (2010). It is the effect of the increased probability of failure of a given institution on the other firms in the economy — both in the financial and in the real sector — and Jeanne and Korinek (2010) attribute it to the debt-asset loop. In the presence of the systemic externality each borrower would find it privately optimal to acquire too much debt, because he does not internalize the spillover effect on others of remaining liquid in a strained episode, relaxing not only his own, but the borrowing constraints of all other borrowers as well.

Governments and central banks are usually able to counter the effects of systemic externalities through bailouts, and of fire sale externalities through purchases of assets or extension of loans (often accompanied by extending the types of assets acceptable as collateral), or, as Minsky would have put it through "protection of assets"⁶⁰. Kocherlakota (2010) admits (in a footnote) that these

⁵⁸ The main result in Korinek (2011a) is that in a decentralized equilibrium bankers insure too little in ex-ante risk markets and engage in excessive fire sales in ex-post asset markets once an adverse shock has materialized. Jeanne and Korinek (2010) show that in the presence of externalities a decentralized equilibrium will arise in which too much debt is acquired.

⁵⁹ The sources of micro-prudential externalities, on the other hand, are asymmetric information and limited liability.

⁶⁰ Regardless of what it is called and how it is implemented, the solution involves the government's assuming private debts on their balance sheets.

measures, through moral hazard, generate a third type of externality: risk externality⁶¹. This externality, which I find is most appropriate to refer to as *risk shifting*, has received quite some attention in the reviewed literature: besides Kocherlakota (2010), it is explored by Shaviro et al. (2011) (credited with “risk externalization”), Keen (2012) (who calls it “bail-out externality”), Weder di Mauro (2010b) (who uses the name “too-big-to-fail externality”), NOU (2011:1) (where it is termed “implicit state guarantee”), IMF (2010) (where it is referred to as a “fiscal externality”) and Acharya et al. (2010). It is used to describe the distortions resulting from the assuredness of governments’ bailing out distressed firms, the collapse of which may have systemic impact, or, alternatively the cost of the desire to protect creditors. This “debt guarantee” (Kocherlakota 2010) provides suppliers of funds with partial insurance against excessive risk, because it reduces the stake they have to fear losing, without affecting the upside earning potential, prompting them to accept lower yields⁶² on their investments for a given amount of risk. As a result, financial institutions are able to finance high-risk, high-return investments at low costs. Thus, the risk externality leads to risk’s being underpriced, so that, in effect, the virtual guarantee of being bailed out is nothing short of a subsidy to the financial sector. The incentives of financial firms are, therefore, distorted (from social point of view, not theirs) towards excessively risky and inefficient investment⁶³, because, drawing on the debt guarantee when choosing their allocation, they fail to internalize the risk their investment decisions impose on others in the financial sector and elsewhere in the economy, in effect shifting risk to these third parties. The debt guarantee also provides scope for moral hazard: Shaviro et al. (2011) go as far as to claim that the prospect of bailouts in the face of major economic setbacks may cause financial firms to prefer large prospective calamities to small ones. Weder di Mauro (2010b) shares this concern: “[w]hile a car driver does not have an incentive to be involved in a big accident, a financial institution does” (Weder di Mauro 2010b:1). It must be noted that the root of the problem is created by the expectations of a “refund” debt holders hold, not by the expectations of financial institutions themselves. Lastly, according to Kocherlakota (2010), financial regulators currently try to combat the ways the risk externality is produced and not its amount.

Keen (2012) identifies a fourth type: a failure externality, which is the social cost arising in the aftermath of a decision to let a systemic institution fail. Its expected cost depends on the probability that a financial entity falls in distress and the wider impact its collapse will have. Obtaining a good estimate of the cost is difficult for two reasons. First, the recent episode of distress showed that there is limited, if any at all, possibility of assessing the impact of an institution’s failing until after it has failed. Second, the determinants of large output losses following in the wake of financial crises must be identified and their workings understood. Like the bailout externality, the failure externality also leads to “subsidized borrowing” but for a different reason: institutions fail to internalize the benefits of higher capital ratios (alternatively lower leverage) which take form in lower risk of own failure and losses accruing to others from trouble in the financial markets or the wider economy.

⁶¹ As I argue in a later section not only do various arrangements allow the financial sector to grow, but financial institutions will exert effort in trying to become systemic. I call this tendency *systemicness bias*. It exacerbates this type of externality.

⁶² For instance, Ueda and Weder di Mauro (2011) estimate the interest rate differential in the presence of bailout guarantees to vary between 10 and 50 bps, with an average of about 20 bps.

⁶³ “Inefficient” is to be understood as “not allocated to uses where it is most productive on the margin”. The misallocation of capital by itself does not give rise to the collective mistakes which result in financial crises, but it means that society loses much more than necessary from them: for instance, Korinek (2011a) formally shows that the very prospect of a bailout exacerbates the suboptimal results to which he arrives in the decentralized scenario of his model, because it induces financial firms to take greater risks than optimal should government intervention be absent, so that their liquidity position after the realization of a shock is below what is privately optimal, and by implication (based on his model) even further below the social optimum.

8.2. Other externalities

The financial sector imposed also other externalities, which have received less attention, probably because they did not involve huge sums of taxpayers' money's being funnelled into the pockets of the financial sector. Nevertheless, as NOU (2011:1) argues, they can have quite serious consequences for the normal functioning of the economy, so they should be taken into account. More specifically, during the crisis many financial institutions had troubles in fulfilling some of their core functions like payment intermediation and others. This had quite some negative impact on consumers and the real economy, which luckily did not go beyond being inconveniences, this time. It is therefore essential for the customers' sake that any changes aimed at dealing with the challenges and weaknesses exposed by the recent crisis also ensure that the trust in financial institutions is retained, at least as far as essential and necessary functions are concerned, so that they are able to meet their obligations at any time. This is important, because many actors, especially smaller ones, hold small buffers to meet freezes in the payment system.

9. Application of Pigouvian Taxation to the Financial Sector – General Considerations

9.1. Pigou's idea

Arthur Pigou developed the idea of using taxes (which eventually came to be known as Pigo(u)vian taxes) as a means of equalizing private with social (marginal) costs, so that externalities are internalized, resulting in a Pareto-superior allocation. His idea has been applied to address both negative, primarily environmental ones, and positive externalities, such as the provision of public goods. In the aftermath of the recent crisis, Pigouvian taxes have been identified as a form of macroprudential regulation in, for instance, Korinek (2011a).

Their theoretical appeal notwithstanding, the practical implementation of Pigouvian taxes is problematic for an array of reasons. First, a major hindrance pointed out already by Pigou himself, is the "knowledge problem": what is the exact amount of externality which has to be corrected for. Second, the unfortunate practice of lobbying by interest groups can affect the tax base, its rate, exemptions, etc. and in the end, the resulting allocation may turn out to be Pareto-inferior to the one one was trying to correct. Third, the substitution effect, which is a corrective tax's main goal, may induce undesirable repercussions, that is, a changed allocation, elsewhere – this may be the case if utility is not separable in its arguments, because then the marginal rate of substitution between any two arguments might depend on at least a third one. Fourth, it is the need for coordination and cooperation among various jurisdictions which is of prime importance in today's globalized world, because without those, a corrective (and in general any other) tax would mostly be avoided to an extent which renders it ineffective.

9.2. How (dis)similar are environmental and "financial pollution"?

First I must specify that I use "financial pollution" as a metaphor for the realization of systemic risk. I draw a parallel to environmental pollution, because in both cases the problem is negative externalities. Since Pigouvian taxation has been applied to deal with the latter, comparing the two would reveal the scope for implementing the same or similar measures to address the externalities in the financial sector. The dominant view found in the literature is that while financial externalities differ little from their environmental counterparts in theory, they possess some peculiar features, which do not allow the direct transfer of solutions in practice. Shaviro et al. (2011) argue that these traits make the implementation of Pigouvian taxation harder than is the case with environmental pollution, while Kocherlakota (2010), in my understanding, argues that there are cases where a direct transfer of solutions is possible, for instance when dealing with the risk externality.

To begin with, in the case of financial externalities there is no distinct output on which to impose the tax, because with many aspects of financial institutions (liquidity, solvency, riskiness, complexity, size of market share, interconnectedness) as a potential source thereof, expected harm cannot be simply aggregated over a single activity⁶⁴. This is often the case for environmental externalities, too. A crucial difference, however, is that for a financial firm *both* its activities *and* the various aspects of its relations with other firms matter. Still, even if some measures of the exerted externalities can be derived, Kocherlakota (2010) argues that the optimal balance between them will depend on a lot of firm-specific information, which is (often legally) hard to obtain⁶⁵. This, together with the necessity of devising a cost minimization procedure for each and every case and each and every institution, makes the calculation of an actuarially fair risk-adjusted fee practically impossible — a point shared by Shaviro et al. (2011) — and hence it may fail to affect the incentives of the firm. In other words, even a well-designed tax may not be able to completely eliminate the incentives to take excessive risk (i.e. “pollute”); certain authors — Acharya et al. (2010) and Weder di Mauro and Klüh (2010) — even argue that it *will* not be able. This is also often the case with environmental pollution.

One crucial difference, pointed out by Weder di Mauro (2010b), is that the amount of systemic risk is endogenous to the reaction of the public sector: the bigger the accident in the financial sector, the higher the chances that there will be a bailout.

9.3. How to measure the amount of the externality?

In the literature there seems to be agreement that the difficulty of measuring the amount of the externality and the contribution by each actor is greater with financial externalities than with environmental ones. Nevertheless, concrete proposals have been put forward. The one by Shin (2010) has already been presented in subsection 5.2.

Korinek (2011a) develops an indicator he calls “externality pricing kernel” which captures the state contingent magnitude of systemic externalities due to risky investments and against which their price can be derived. It is defined as the difference⁶⁶ between the social and private costs of a payoff in a particular state, which, equivalently, captures the uninternalized social cost of financially constrained suppliers’ of credit having to make a unit’s payment in that particular state. It is zero in unconstrained states and positive in constrained ones. A security which co-varies positively with the externality kernel imposes a negative externality, while one which co-varies negatively imposes a positive externality.

Kocherlakota (2010) comes with another proposal. He argues that a market-based mechanism of quantifying risk (i.e. the externality) is superior to regulatory monitoring, because the latter suffers from informational asymmetry operating along two dimensions: the monitoring required may be technically infeasible or prohibitively expensive, and the institution in question’s being very likely to be aware of what supervisors are doing, it may try to draw some benefit from pretending to be of a particular type through signalling and mimicking. Kocherlakota envisions the government’s issuing a “rescue bond” for each relevant institution, which pays a variable coupon — zero most of the time, but potentially quite a sum in the rare events of crises — of some fixed fraction of the actual transfers from taxpayers to the institution and/or its stakeholders. In a perfectly functioning market the price of the bond will be equal to the fraction of the expected discounted value of the transfers to the institution’s stakeholders, thus reflecting the market’s judgement about the institution’s risk

⁶⁴ According to some authors, this is the likely explanation for the existence of separate regulations for each of those various aspects.

⁶⁵ Kocherlakota (2010) is also dismissive of the use of information acquired through indirect channels, because its reliability is likely to be questionable.

⁶⁶ In the model of the paper it is the difference between the private valuation of liquidity by bankers and the social planner’s valuation of liquidity. The decentralized and the social planner allocations differ because the respective valuations of liquidity in constrained states differ.

profile. The authorities can, therefore, charge the institution a tax or a fee equal to the reciprocal of the fixed fraction times the price of the bond. This approach has two advantages: first, it could be used for a variety of institutions, not just banks; second, at least in theory, the authorities will be spared pondering over the systemic importance of one or another institution, because the information will be provided by the market⁶⁷ through the price of rescue bonds. The market approach faces challenges in practice, though. For instance, the markets for rescue bonds may prove thin and illiquid, in which case other indicators will have to be employed to obtain an appropriate measure of the externality⁶⁸. Further, how do we know — if we can at all — that the market will price risk correctly *this time*?

The proposal by Acharya et al. (2010) is quite similar to the previous one, but they envisage the purchase of contingent capital insurance for systemic risk by each firm, where the price paid will reflect the relative danger the entity in question poses to the financial sector.

9.4. What can the various proposals to employ corrective taxation achieve?

Corrective taxation would ideally equalize private marginal cost with social marginal cost, forcing the “producers” of the externality to internalize it and hence bring its amount to the socially optimal level. In the literature it is pointed out that the presence of financial frictions, due to information’s being imperfect and costly to obtain, imposes a constraint, so that the realized Pareto-optimum will not be the unconstrained first-best one. What concretely are appropriately designed Pigouvian taxes likely to achieve?

First and foremost, corrective taxes will shift some of the costs of financial crises back to the financial sector from third parties. The latter have to shoulder the difference between the often difficult to measure overall cost of a financial crisis and the direct fiscal cost of providing support to distressed institutions and economic stimuli. Kocherlakota (2010) argues that if a financial institution is aware that it faces this cost, its private costs of investment will be equalized to the social costs of investment, hence the investment choice will be efficient, translating in efficient levels of firm-specific factors like capital, liquidity and remuneration. Closely related to this is another outcome: to the extent that taxes and fees will force the participants in the financial sector to internalize the social costs of their engaging in risky activities, these instruments of macroprudential policy can help improve the pricing of risk, thereby reducing riskiness of positions.

Further, the optimal structure of funding can also be addressed by Pigouvian taxes. The observed tendencies of financial institutions to become overleveraged over time and to rely heavily on short-term debt can be countered by a corrective tax equating the private marginal cost of borrowing with the social marginal cost of borrowing. More specifically, Shin (2010) argues that a Pigouvian tax can be imposed on non-core liabilities, with the goal of preventing excessive use of this source of financing, while *The Mirrlees Review* (2011) mentions the use of taxes as one of the means of combating the excessive use of short-term liabilities (that is, the pecuniary externality identified above). The failure externality, identified in Keen (2011), can also be fought against with taxes. Huang and Ratnovski (2009) argue that the problem of suboptimal use of senior wholesale funds by banks can be solved through a Pigouvian tax on the use of short-term wholesale funds such as collateralized repos, but warn, though, that such a tax has its limitations, namely that it is not as broad as a systemic risk tax, even though it shares some intuition with it.

⁶⁷ Besides Kocherlakota (2010), Acharya et al. (2010), and Jeanne and Korinek (2010) also assume that private agents are capable of correctly pricing risk. The recent crisis showed, however, that such an assumption is not empirically substantiated.

⁶⁸ Kocherlakota notes that even if they are imperfect, market measures still contain valuable information which can be made use of in the supervisory process, for example when it comes to determining the appropriate discount rate.

Weder di Mauro (2010b) proposes the use of a systemic risk charge as a means of addressing the too-big-to-fail externality. It should ideally try to tax away the incentive to engage in excessive risk taking because of lower borrowing costs resulting from the bailout guarantee. The same view is expressed by Kocherlakota (2010:1) who believes that “[a] well-designed tax system can entirely eliminate the risk externality generated by inevitable government bailout”, and Shaviro et al. (2011) who argue that even if corrective taxes or fees are not precise for each case, they can still neutralize the effect of the subsidy, provided that one gets them right on average. Closely related is the argument by Acharya et al. (2010) that Pigouvian taxes can be used to incentivize firms to become less systemic, that is, as a remedy against what I term systemicness bias. Unlike the previous authors, however, but together with IMF (2010) and Weder di Mauro and Klüh (2010) they do not believe that a complete elimination of excessive risk is possible – in equilibrium some institutions will still find it optimal to engage in risky behaviour, in spite of the higher taxes they will be paying.

Korinek (2011a) argues that a state-contingent tax on the issuance of Arrow-Debreu securities can equate the private and social valuations of liquidity across all states of nature, alleviating the systemic externality. The result can be extended to more complex securities, since in theory they can be represented as a collection of weighted Arrow-Debreu securities.

9.5. What is the implementable optimal corrective tax likely to look like?

In theory, which Shaviro et al. (2011) define as absence of administrative, implementation and political economy issues, a corrective tax is conceivable for any thinkable externality, and it is bound to be very complex. The practically implementable options, however, are much fewer and simpler. The pool of relevant realizations is further limited by the different strengths and weaknesses of each proposal. Some criteria for comparison are necessary. One set of such can be found in Weder di Mauro (2010b) and consists of four desirable characteristics which a tool should possess to be effective at internalizing systemic risk. First, it should ensure private incentive compatibility, that is, reduce private incentives to create or increase systemic risk. Second, it should induce public incentive compatibility, that is, reduce authorities’ incentives to use forbearance and offer bail-outs⁶⁹. Third, it should provide transparency to the supervisor, the markets, and the public. Fourth, without doubt the most difficult one to realize in practice, it should be easy to enforce and difficult to circumvent.

Unlike their counterparts on environmental pollution, almost any Pigouvian taxes on “financial pollution” will be institution-specific and state-specific, and will have to reflect concerns about multiple risk-related attributes, making them potentially difficult to calculate and implement. There seems to be unanimity in the literature that the tax rate should vary non-linearly with the size of the externality. Kocherlakota (2010) further argues that an optimal tax based on the measure he proposes will vary over time, because financial institutions will take into account the impact of new information when making their decisions. Both he and Korinek (2011a) argue that financial institutions can act or use securities to reduce risk, so that the optimal tax can be negative, to encourage such practices. Lastly, all else equal, the tax rate should be higher for higher leverage.

9.6. How is the optimal Pigouvian tax likely to depend on the parameters in the economy?

As part of their study of the effect of taxes on credit booms and busts, Jeanne and Korinek (2010) provide sensitivity analysis which is applicable outside of their model as well. First, they point out that attempts to deal with financial sector externalities can come into conflict with other policy objectives, such a stimulating demand. More specifically, lower interest rates warrant higher

⁶⁹ This means that a solution to the time-inconsistency problem has to be found. According to Weder di Mauro and Klüh (2010) there is very real danger that the incentive problems of government authorities vis-à-vis the financial sector may remain unaddressed.

corrective tax rates to prevent excessive debt accumulation. This may well be welfare-reducing, because certain loans, which would constitute Pareto improvement absent the tax, will not be taken.

Further, Jeanne and Korinek (2010) point out that greater financial liberalization and financial engineering warrant tighter macroprudential regulation over a significant part of the parameter space implying that the optimal tax rate should increase with the ability of private agents to borrow, because the more that can be borrowed against collateral, the greater the potential amplification effects when borrowing constraints eventually bind. They also determine whether prudential taxation should be transitory or permanent: the lower the level, the shorter the duration of necessary application.

Lastly, Jeanne and Korinek (2010) find that the optimal tax is not very sensitive to changes in the size and probability of the underlying shock which they attribute to the endogenous response of borrowers, who increase their precautionary savings in the face of increased riskiness in the economy.

10. Incentives

The individual behaviour of the actors in the financial markets — engaging in profit (rent) maximization — was in accordance with theory. According to Shaviro et al. (2011), the realized socially suboptimal allocation is the result of their facing the socially wrong incentives: the framework, within which they had to operate, exposed them to significant moral hazard. Specifically, it did not pay off to be prudent. As a consequence, the actors in the financial markets acquired what eventually turned out to be an unmanageable amount of leverage, took excessive risks, while the financial sector grew in size beyond the needs of the economy.

Incentives and behaviour being central in game theory, it seems natural to present my interpretation of imprudence in game-theoretic terms. The existence of externalities gives rise to what is known as the “first strategic tension” (Watson 2008): the conflict between an individually rational action — seeking and reaping rents — and a collectively rational action — not jeopardizing the stability of the financial system. The interactions between the actors in the financial sector can be represented as an n-player Prisoner’s dilemma. If they are modelled as a static or finite game the only equilibrium involves excessive rent-seeking. If, on the other hand, the interactions are modelled as an infinitely repeated game — which seems to be the most adequate representation of reality — cooperation around the Pareto-superior equilibrium is possible, provided that the players do not discount the future very heavily. However, the tremendous shortening of horizons in financial institutions in the recent decades is evidence that the future has been discounted heavily⁷⁰. This, in turn, would imply that the only stable equilibrium in a decentralized setting is the egoistic one. The financial crisis is an empirical confirmation that this, indeed, is the case.

I find the essentially same conclusion in Korinek (2011a) where the effects of market discipline are discussed. The analysis suggests that unless externalities are dealt with, markets will punish prudent banks behaving socially responsibly and reward banks taking socially excessive amount of risk.

There seems to be agreement in the literature that the way the crisis was resolved has increased the scope for moral hazard — it is argued that the extensive interventions have strengthened the expectations that authorities will step in to avert another crisis, providing support both to the market and to specific institutions; Weder di Mauro and Klüh (2010) argue that as a result of the operation of

⁷⁰ This is under the assumption that preferences are time-consistent. Short-termism can alternatively be rationalized by using time-inconsistent preferences, such as present-biased ones, the existence of which has been confirmed in various empirical studies in behavioural economics.

this “ratchet effect”, as they call it, authorities place themselves in an increasingly disadvantageous position vis-à-vis the financial sector. While such expectations need not be fulfilled — consider the failure of Lehman Brothers — a refusal by the authorities to provide assistance will be dismissed as time-inconsistent by rational actors. The scope for moral hazard has been further strengthened by the lack of any serious attempts to recoup a sizeable amount of the rents which accrued to the “winners” of the financial crisis.

10.1. Systemicness bias

Rent seeking has two dimensions. The first, which is always present, is simply the very process of reaping the rents. The second one, which need not be always present, consists of conscious efforts to create rents⁷¹. Financial institutions have been quite successful along the second dimension, but this has received unduly little attention in the literature.

The ability to obtain rents is known as “market power” and every profit-maximizing entity will naturally seek ways to gain, increase or retain it. There is, however, a difference: the exercise of market power by a financial institution, as rule, confers on it the status of “too-big-to-fail”⁷² or “systemic”; only in exceptional cases does this apply to non-financial firms. Rational profit-maximizing financial institutions, being well aware of the advantages associated with this status, will therefore *consciously* attempt to become too big to fail, or pretend to be. I call this a *systemicness bias*. It has at least two unpleasant effects. First, the more systemic an institution is, the more distorted its incentives will be. Second, the market structure will be distorted towards one favouring big, systemic institutions, which even in the absence of financial troubles will be able to enjoy traditional rents by offering less favourable conditions to their customers.

The literature would appear to have accorded rather little attention to this bias: Ueda and Weder di Mauro (2010:106) explicitly state that the persistent nature of explicit and implicit subsidies indeed “creates perverse incentives to become systemically relevant by becoming larger and more complex”. They also attempt to quantify these incentives by using as proxies either funding cost advantages enjoyed by institutions deemed likely to receive government support — their study shows that expectations of intervention were high already before the crisis — or what they call “the excess bailout benefits to TBTF [too big to fail] institutions in crisis” (Ueda and Weder di Mauro 2010:107). The two authors argue that a well-designed tax will offset the too-big-to-fail subsidy, thereby helping address the distortions it causes; this view is shared by IMF (2010) as well. Acharya et al. (2010) point out that corrective taxation, *in addition* to forcing financial firms to internalize the costs of their actions, should also provide them with incentives to become less systemic — clearly this is more of a side matter in their analysis, while Weder di Mauro and Klüh (2010) just mention that an appropriate regulation framework for the financial sector should make the acquisition of systemic relevance a costly matter. On the other hand, even though none of the mentioned references deals primarily with the systemicness bias, IMF (2010:7) identifies “measures...reduc[ing the] incentives for financial institutions to become too systemically important to be permitted to fail” as one of the two key objectives which should guide the work towards efficiently dealing with failing financial institutions in the future. The dominant view in the literature seems to favour subjecting systemic institutions, especially those of global importance, to a combination of higher pecuniary and more stringent non-pecuniary requirements so that their ability to meet losses is strengthened. In my opinion, such arrangements can also serve as effective counter-measures against the systemicness

⁷¹ One example would be a natural monopolist — it enjoys rents which are present even without its conscious efforts to create them.

⁷² In general, an entity is considered too-big-to-fail/systemic if the costs to the wider economy of its being let fail — known as “costs of liquidation” — are deemed prohibitively high. These prohibitive costs of liquidation are the grounds on which authorities try to sell to the public the transfer of funds to the firms in question, in effect putting a limit to the downside risk they bear. Changing circumstances may and often will cause the length of the list of too-big-to-fail’s to vary.

bias. Further, the distortion of incentives and market structure is another justification for corrective taxes' and fees' being non-linear.

10.2. Advantages of being too big to fail

Here I briefly mention how a financial institution can benefit from its too-big-to-fail status. The issue, unlike its cause, the systemicness bias, has received considerable attention in the post-crisis literature. A too-big-to-fail institution operates with an implicit state guarantee which confers two advantages identified in the literature. First, it acts as a de facto subsidy lowering financing costs of the particular entity and allowing the acquisition of leverage, more so the more systemically relevant the financial firm is. Ueda and Weder di Mauro (2010) estimate the value of this subsidy to entities too big to fail to be in the range of 10 to 50 bps, averaging about 20 bps. Second, it acts as an insurance which the institution enjoys, but for which it does not pay adequate or any premia, providing a safeguard against greater risks. Further, these two advantages combine to provide a competitive advantage to too-big-to-fails over non-systemic institutions, allowing them to enjoy additional rents from market power, which they, according to Ueda and Weder di Mauro (2010) can in turn use to further increase the funding cost differential vis-à-vis other entities, thereby exacerbating the effect of the implicit subsidy they enjoy.

11. Debt Bias

I use the term "debt bias" to refer to the systematic preference for debt over other forms of finance, which leads to a tendency of acquiring eventually unsustainable levels of leverage. It is a phenomenon working on the aggregate level which materializes over the medium term (10-15 years). Another interpretation is offered by Jeanne and Korinek (2010), who model it as an externality arising together with asset price inflation in a decentralized setting because the failure of borrowers to internalize spillover effects of financial amplification causes them to acquire too much debt in good times.

One point where there is apparent agreement between the mainstream and non-mainstream economic literature is that systemic risk, both within and outside of the financial sector, increases with the debt level. Keen (2012), NOU (2011:1), Shin (2010), Shaviro et al. (2011) all point out that a major deficiency of the regulatory framework, which both contributed to and was revealed by the recent crisis, is the failure to affect incentives to acquire too much debt in a socially Pareto-superior direction.

Why is it important to prevent the accumulation of excessive debt? Based on the reviewed literature, it appears to me that the mainstream stresses the dangers of rapid deleveraging of financial institutions because of their role in the financial amplification cycle. On the other hand, the main peril of excessive debt I find in the non-mainstream literature is its role in the formation of bubbles. Both strands accord also attention to the fact that willingness to engage in risky activities increases with leverage, especially if the authorities are expected to validate the decisions to incur debts (which they are more likely to do the higher the liabilities).

11.1. Tax distortions and incentives to acquire debt

Various authors have made the observation that many current tax systems' deviation from the neutrality principle may have adverse effects on financial stability, because they favour the build-up of financial imbalances – indeed IMF (2010:3) puts it bluntly: "...current tax distortions that run counter to regulatory and stability objectives". This "tax bias", as Shaviro et al. (2011) and Keen

(2011) call it, is the result of the asymmetric tax treatment of debt and equity favouring the former⁷³, which translates in systematic preference for debt finance⁷⁴, a fact that “at a minimum, sits oddly with regulatory measures intended to do exactly the opposite⁷⁵” (Keen and de Mooij 2012:3). Financial engineering further magnified the distortions through the development of what is known as hybrid instruments – financial products which appeared to offer the opportunity of having one’s cake and eating, too, by counting as debt for tax purposes and as equity for regulatory ones, thus “conflicting with regulatory objectives” (IMF 2010:62).

The limited recognition of the distortions which taxes cause on the financing structure of financial firms translates into a lack of a clear consensus in theory on the role of tax asymmetries on the capital structure. Keen and de Mooij (2012) argue, therefore, that the issue is an empirical one, but they, together with Keen (2011), report that virtually no research on the topic has been carried out. The former article presents two alternative views on why the tax asymmetry can affect the capital structure of financial firms differently. According to the first one, taxes do not have significant impact on the capital structure of financial firms, especially banks, because their financing decisions are limited by the regulatory framework in place. This view, however, ignores the fact that banks often keep buffers of equity in excess of capital requirements, which leaves scope for tax effects on leverage and capital ratios after all. Still, when their model is put to the test, Keen and de Mooij (2012) do not find that capital requirements make banks more responsive to corporate tax than non-financial firms. The other view is on the opposite extreme: the capital structures of financial firms are more tax-sensitive, because the nature of their business allows them to exploit the tax asymmetry by employing strategies which are not available to non-financial firms, such as use of hybrid instruments and taking advantage of both implicit (too-big-to-fail) and, if applicable, explicit (deposit insurance) guarantees on their liabilities. This second view is reflected in one of the predictions of the model of Keen and de Mooij (2012), namely that for a given level of hybrid use an increase in the corporate tax rate will cause banks to increase their total borrowing. However, for a given total leverage the use of hybrid securities is independent of the tax rate. The authors explain this “surprising result⁷⁶” with financial firms’ basing their decisions on the *total* amount of debt, regardless of instrument, and hybrid securities provide an opportunity for regulatory arbitrage, so that there is a “pecking order” (Keen and de Mooij 2012) effect at work: the full amount of hybrid securities permitted by regulation should be exhausted first, only then will other forms of debt be considered. The reason is the dual nature of hybrid securities: they both confer the same tax advantage as other forms of debt and count as equity for meeting capital requirements. The empirical findings presented in the article support this result. Thus, while the use of hybrid securities is insensitive to taxation (although it still increases somewhat with the tax rate), the ability to adjust their use makes total leverage more responsive to taxation.

11.2. Policy response

The results in Keen and de Mooij (2012) suggest that the favourable treatment of debt finance for corporate tax purposes does lead banks to acquire more leverage than what would be the case otherwise, regardless of the mandated capital requirements they have to adhere to. So, elimination of the tax bias would potentially increase bank capital both in the short and in the long run. However,

⁷³ For instance, interest payments on debt are deductible against the corporate income tax, while cost of, and payments to equity normally are not. The practice of deducting debts for taxation purposes has received a lot of attention, especially in the case of deductibility of home mortgages; the latter is regarded as a main factor behind housing bubbles.

⁷⁴ The preference for debt over equity is by no means limited to financial firms. It is their impact on the wider economy which makes the issue acute.

⁷⁵ This is not a universal phenomenon, though: NOU (2011:1) credits the rather neutral tax system in Norway for the equal treatment of debt and equity financing.

⁷⁶ The surprise is, according to Keen and de Mooij (2012), that hybrid securities are in practice instruments for tax avoidance, so one would expect their use to be strongly sensitive to taxation.

doing so in the case of the systemic banks is problematic, because of the relative insensitivity of their leverage ratios to either of taxation and capital requirements. This implies that far-reaching reforms are needed if they are to be induced to change the composition of their capital structures in a socially superior way.

Keen and de Mooij (2012) offer as one possible solution the use of explicit constraints on leverage ratios – a measure that has found its way into the Basel III framework. They argue that in the case at hand, changes in regulation are better posed to deal with the aspects of capital structures most closely associated with the externalities caused by excessive leverage, and hence it is the more effective strategy to pursue.

Another way of reducing incentives to acquire excessive leverage on the micro-level is through reforms in the corporate income tax, so that debt and equity receive equal treatment. More specifically, this could be implemented through reduction or elimination of interest deductibility altogether, which is part of what IMF (2010) and *The Mirrlees Review* (2011) refer to as a “Comprehensive Business Income Tax”; see 17.3.3. in the latter source for more on the topic. However, the former reference cautions that the transition to this new arrangement will entail significant problems, so that, so far, only an in-between solution, called “thin capitalization rules”, whereby interest deduction is not applicable beyond certain threshold levels for debt ratios and interest outlays, has been implemented. Alternatively, allowance for corporate equity with a corresponding reduction of interest deductibility may be introduced, so that revenue is conserved. More generally, IMF (2010) sees a fundamental reform in corporate income taxation as a part of the overall attempt to improve the taxation of the financial sector, arguing that while such profound changes will involve difficulties, “the payoff to reducing a fundamental bias to excess leverage could be substantial” (IMF 2010:63).

A third option is proposed by Matheson (2011). He envisages a tax levied on the debt in a bank’s balance sheet. Liabilities net of equity and insured deposits will comprise its base. The schedule should be progressive according to the type of institution, so that larger and more systemically important banks are taxed more heavily⁷⁷.

Based on their view of the debt bias as an externality, Jeanne and Korinek (2010) show that a Pigouvian tax on borrowing may help deal with the issue. However, this proposal of theirs aims also at mitigating the effects of swings in asset prices, so it will be presented in the next section.

11.3. The interaction of leverage and risk

Keen and de Mooij (2012) offer a brief discussion on the topic, because the focus of their article is on leverage. Still, the issue is important, because the riskiness of a position can affect leverage decisions or vice versa, for instance through gambling for resurrection. In addition, taxes levied on liabilities may therefore have effects on risk strategies employed. The model of the paper implies that, absent agency costs and for a given amount of hybrid instruments used, increased riskiness of asset positions, measured by the fraction of total investment put into risky projects, should lead to reduction of total leverage. On the other hand, increased riskiness of asset positions would lead to increased use of hybrid instruments at the expense of other debt instruments. In the absence of agency costs and the unavailability of hybrid instruments a rise in the corporate tax rate would lead to a reduction in the proportion of assets invested into risky projects. When Keen and de Mooij (2012) put the model to the test, they find that neither the size, nor the significance of the “tax bias” are affected by the inclusion of risk; the explanation they offer for this finding is that considerations other than the central in their model links with regulatory capital may shape the interaction between risk and leverage. In line with the theoretical predictions the empirical results show the existence of

⁷⁷ This proposal is virtually identical to the one by Shin (2010) mentioned previously. However, each aims at a different goal. The proposal by Matheson (2011) pursues sustainable leverage, whereas the one in Shin (2010) is an attempt to limit excessive engagement in risky activities. In my opinion, the acquisition of excessive leverage in view of the recent crisis (at least) may be considered a risky activity, so the goal of the two solutions may seem to be the same after all.

statistically significant positive correlation between the riskiness of an institution's assets and its use of hybrid instruments. Further, the significant reduction in leverage ratios due to stricter capital requirements is amplified by the riskiness of the asset positions. Again in line with theoretical predictions, higher corporate tax rates will be met with increased leverage and reduced riskiness of positions; the latter may reduce the cost of violating the capital requirement.

12. Booms and Busts in Credit and Asset Prices

The recent crisis was the latest confirmation of the existence of systemic interaction between debt accumulation and asset prices and its importance in many economic settings, of primary interest among which is its contribution to magnifying the effects of booms and busts. In spite of the relevance of the topic, Jeanne and Korinek (2010) acknowledge a paucity of welfare analyses of optimal policies addressing the fluctuations in debt and asset prices.

In the article they argue that the debt-asset loop produces systemic externalities as a result of which both credit flows and asset prices oscillate with greater amplitude. Within their model a counter-cyclical Pigouvian tax on borrowing may increase welfare by inducing each borrower to internalize the cost of relaxing others' borrowing constraints as a result of his own actions. The time path of the tax should be characterized by gradualism: even if the debt level may be low, it should increase gradually with the economy's vulnerability to a new credit crunch. The tax would ideally be non-linear and, importantly, differentiated so that its rate should be adjusted for each sector's susceptibility to sudden precipitous declines in asset prices and availability of credit.

Jeanne and Korinek (2010) find that the ability of a counter-cyclical tax on borrowing to produce Pareto improvement is not affected by three extensions to their basic model: a shock to income, a shock to the availability of credit⁷⁸ and issuance of long-term debt or equity by the borrowers. More specifically, the size of the systemic externality and hence of the optimal policy measures attempting to internalize it depend on the extent of amplification when borrowing constraints bind rather than on the source of shocks. Longer maturity of debt mitigates the effects of debt deflation⁷⁹ when borrowing constraints bind by making the economy more resilient, because it reduces the scope for multiple equilibria: there is a smaller fraction of debt coming due in every period, which also protects borrowers against self-fulfilling panics. The need for prudential taxation of debt is not diminished by the ability to issue equity, because its better risk-sharing properties notwithstanding, it is not used to reduce risk in equilibrium, but to increase consumption, so that borrowers end up with the same level of debt.

13. Information Issues

The operational framework of the financial markets, especially the one produced by the process of financialization, does not allow for another of the prerequisites for the First Main Theorem of Welfare Economics — that of complete and symmetric information — to hold. As a result, there arose informational rents which exacerbated the scope of moral hazard; needless to say that the actors in the financial sector did not encounter significant difficulties in reaping them. Further, the

⁷⁸ Within the framework of Minsky, on the other hand, the nature of the shock would make a difference, because a shock to income affects both hedge and speculative borrowers, while a shock to the availability of credit affects speculative borrowers only.

⁷⁹ This result is in line with experience: it is short-term debt and the necessity to roll it over which is the problem. See Minsky (1982; 1986) and NOU (2011:1) for a detailed discussion on the advantages of long-term debt.

gap between private and social costs in a deregulated market is indicative of prices' conveying the wrong information which, in turn, resulted in a socially suboptimal allocation. As a result of the crisis, the issues of transparency and openness have finally received due attention.

Asymmetric and incomplete information on the micro-level have been thoroughly studied in the literature on corporate finance. However, the forms and implications they have on the macro-level are different. More specifically, important aspects of financial institutions are better known to them than to regulators. This asymmetry is the main argument for the virtually universal proposition that differential treatment is necessary, should taxes be used as a corrective measure.

13.1. Gains from reducing the scope of information distortion

According to NOU (2011:1), the reduction of the informational advantage of financial institutions regarding their financing structure and liquidity risk will benefit their counterparties by allowing the latter to make better-informed choices. This will also likely reduce the scale of liquidity freezes in the future. Informational disadvantages contribute to increased uncertainty, which then can by itself contribute to increased risk. Addressing the issue can, therefore, help to avoid excessive risk. Lastly, the efficacy of the actions undertaken by regulators and supervisors will be higher the less intransparent the environment is in which they operate.

Uncertain market conditions can create appropriate ambience for market manipulation, which can destabilize the financial system and thence through self-fulfilling prophecies the real economy. The issue is especially relevant in view of the scandals in the last couple of years concerning the manipulation of various reference rates (e.g. LIBOR) and markets (e.g. foreign exchange, gold). Addressing the problem can, therefore, contribute to increased stability of the financial system, all else equal. Market manipulation and the closely related issue of fraud have received attention in NOU (2011:1) and in most of the pre-crisis writings. However, no concrete countermeasures have been discussed in the extant literature.

13.2. How could taxes and fees be used to address information issues?

Taxes and fees can be used to diminish the informational rents accruing to financial institutions from the informational advantage they enjoy, thus also reducing the scope of the latter – in effect, financial entities will have to provide more information at their expense, so that their operational costs increase. According to NOU (2011:1), such a scenario is welfare-improving, because the extra expenses will be more than outweighed by the gains to the customers and counterparties from their ability to make better-informed decisions.

The report does not mention certain practical issues, which I believe are important. First and foremost, it must be provided for that the financial institutions do not pass these extra costs to their customers in the form of higher prices or lower returns, at least not all of them. Second, it is important that the information provided by the institutions meets the criteria set by the regulators.

When it comes to issues of compliance, the pro-active perspective of the current composition translates into the preferred solution's being to provide ex-ante incentives to financial institutions to provide information, rather than punishing them ex post for failing to do so. I envisage a relatively simple way of doing this. At the beginning of each financial year, each institution will have to deposit a certain amount with, say, the Financial Supervisory Agency. Its composition can be based on various attributes of the institution such as, for instance, its size and importance on the domestic financial market, whether it operates abroad, its track record (previous delinquencies and the like). The Financial Supervisory Agency will then be given the mandate to verify whether the institution has met the requirements for information provision throughout the financial year. If it has, at the beginning of the following financial year, it will be paid back the amount, with optional compensation for the inflation over the period (or subtraction of the gains from potential deflation). If the Financial Supervisory Agency deems the information provision unsatisfactory, it will retain the amount, and

should be given the opportunity to fine the institution, if it finds that necessary. The financial institution should not be allowed to appeal the decision for fee retention in court, neither the imposition of the fine, but it might be allowed to contend its amount. The idea behind my proposal is that the institutions will have to pay for providing wrong or deficient information which they, as cost-minimizing units, will try to avoid. This, in turn, will give them any incentive to be transparent. To incentivize financial institutions to comply with regulations over the longer term interest may be paid on the amount deposited, whereby compliance over a run of years will be required before any interest pay-outs can start. The interest paid may increase over time based on some formula so that the longer the uninterrupted compliance, the more will accrue to the financial institution. Of course, should it fail to meet the requirements for information provision, it will have to start the whole process all over again.

14. Size of the Financial Sector

Among the lessons to be drawn from the recent crisis is the significant importance of the financial sector's size and growth; for instance Laeven and Valencia (2010b) attribute its larger economic costs compared to previous crises in part to the increase the size of the financial industry. Nevertheless, currently too little attention and efforts are devoted to the issue in the process of overhauling the regulatory and supervisory framework. There is agreement in the post-crisis literature that if, in the future, the financial sector is charged adequately, through regulation and taxation, for the risk its actions pose to the rest of the economy, one likely collateral consequence is that it will become smaller⁸⁰.

14.1. The size of the financial sector from the perspective of a particular country

Any sector comprising a large fraction of a country's economy is a potential source of significant danger to that state, and, through various forms of interconnectedness, also to other ones. The financial sector is no exception – the recent crisis demonstrated that it has large systemic effects for a country's economy and can impose huge economic obligations⁸¹ on the state — Iceland and Ireland are cases in point.

How big a financial sector is too big? Well, when a financial crisis threatens not just the economic stability of a country, and neither the government, nor the other sectors of the economy are able to deal with the consequences of the failure of the financial sector, then the latter is obviously too big.

According to NOU (2011:1), there can be several reasons for why a country's financial sector may become too large. First, this may be the result of high indebtedness or imbalances in the state's economy, making it more vulnerable to increased dangers of financial and economic crises, and exacerbating the severity of downturns. Second, a country's financial sector can become big if its constituent institutions have large exposure to foreign markets; the country's economy becomes thereby more vulnerable to disturbances in other countries than what is the case of exposure

⁸⁰ That is to say its size will be affected even if one has no intention of doing it. This is the case already with, for instance, capital requirements – they impose some limits on size and growth, even though this is more of a side effect, rather than a deliberately pursued goal.

⁸¹ Too big a financial sector can reduce the ability of the government to support it (and with it the economy as whole) during times of distress. This can weaken the trust markets have in the institutions in the financial sector, contributing to instability. It also opens up the possibility for particular agents, for instance holders of credit default swaps, to start betting on particular developments. Something similar has been experienced before during episodes of speculative attacks on various currencies.

through other sectors of the economy. Third, a large financial sector may be the result of deliberate government promotion.

The report points out that the development has not been quite uniform across countries. Besides more traditional countries of financial services, like the USA, UK and Switzerland, the financial sector grew significantly prior to the crisis in Iceland, Ireland, and Cyprus. The amount of capital managed by the financial sector in Norway also grew faster than GDP in the first decade of the current century, but this was mainly due to a rise in extension of traditional loans to households and businesses, not to financial innovation and securitization markets⁸². While such loans have historically been of little risk to the issuer, the increase in supply of funds to borrowers may have destabilizing effects on the macroeconomic level, because it can contribute to the formation of price bubbles⁸³.

14.2. Why the financial sector can easily become large in general

The ability of the financial sector to grow fast is conditioned by the lack of any serious capacity constraints, as financial products can be produced without limit, with derivatives a stark recent example. Rapid expansion of the financial sector's product catalogue is normally accompanied by significant hidden risks, because it results in structural changes the consequences of which will rarely be known at the appropriate time.

As was said before, certain peculiarities of the financial sector, which set it aside from other sectors in the economy, can often be used as a starting point for explaining why it behaves in ways inconsistent with mainstream economic theory. Implication would suggest that some of these oddities may be among the reasons for why the financial sector is able to grow too fast and become too big. And, indeed, this is the case. NOU (2011:1) identifies four main features on which this ability and opportunity rest.

First, the significant informational advantage of financial institutions over their customers allows them to appropriate the resulting rents in the form of higher profits, for instance, through the sale of overpriced products and services.

Second, the implicit state guarantee — a de facto subsidy — enjoyed by many financial institutions enhances their profitability by relaxing their borrowing constraints, making it easier and cheaper⁸⁴ for them to obtain financing. This feature is no triviality: the rating agencies take explicitly into account the probability of and the value of government support to financial institutions and financial markets when they set their ratings. More generally, economic intuition predicts that the value creditors attach to expected support is an increasing function of the probability thereof, that is, the probability of a crisis. The prospect of intervention on behalf of a troubled institution acts further as a positive feedback: big financial firms obtain cheap financing, which enhances their profits and growth. This enables them to become even bigger, get even better borrowing conditions and potentially pose even bigger threat to financial stability. Not everybody agrees on this aspect, though: Shaviro et al. (2011) quote the findings of other (that is, in their reference list) authors, who claim that the financial sector is too large for reasons unconnected with the implicit subsidy it enjoys.

⁸² Nevertheless, according to NOU (2011:1), increased dependence on market finance was also a factor behind the high growth.

⁸³ This is an example of the fact mentioned earlier that stability on the micro-level need not imply stability on the macro-level.

⁸⁴ Even though it is not possible to observe directly to what extent the costs of financing are influenced by expectations of government support measures, measurement attempts have been undertaken: for instance, NOU (2011:1) cites Andrew Haldane's 2010 study "The \$100 Billion Question" according to which the yearly subsidy to the largest five banks in the UK amounts approximately to their profits. The report estimates the value of the implicit state guarantee to the Norwegian financial sector to be around five milliard NOK.

Third, as part of its favourable tax treatment, the financial sector is undertaxed compared to other sectors. While there are no clear-cut results on the net effect of the mispricing on the various customer groups, there is evidence in studies that the exemption makes the sector appear too big, because it can skew consumption away from other goods and services to financial products – as Cottarelli (2010:76) has put it “...a tendency for the financial sector, purely for tax reasons, to be too large—or too fat”.

Fourth, the financial sector disposes of enormous resources and enormous political and economic influence. This may have played a role in the unfortunate development in the pre-crisis years, when the operational framework of the sector was formed in a way which favoured its expansion and profitability at the expense of other considerations.

15. The Bounds of Systemicness

As I have argued before, economic intuition implies that financial institutions will tend to put effort into acquiring a too-big-to-fail status, or appear to have it, a phenomenon which I call *systemicness bias*. Now, this raises the question how big a particular financial entity has to be so that the state intervenes on its behalf, that is, we need some measure of systemic importance. Such has so far been difficult to agree on. A lot of models have been proposed in the ongoing discussion internationally. It is a different question whether there are theoretical justifications for the existence of systemic institutions. As a friend of mine once said “if you are too big to fail, then you are too big to exist”.

In general, an institution can be considered systemic if it is behind a significant portion of the credit supply, and/or if it is an important counterparty for other financial institutions – in economic terms this means that it bears *a lot of* default and counterparty risk. Neither bigness, nor complexity is necessary for an entity to be (considered) systemic⁸⁵. Further, in times of trouble more institutions may have to be considered systemic than what is the case in normal times; the recent crisis is an extreme example of this fact – *any* institution was considered systemic when it came to providing support measures. According to Weder di Mauro and Klüh (2010) the latter can in part be attributed to the lack of appropriate resolution mechanisms for entities of systemic importance.

In the aftermath of the crisis, the question has been raised whether special regulation should be put in place for systemic institutions because of the challenges they pose, such as the virtual necessity of bailing them out, or the insensitivity of their leverage ratios to taxation and regulation, as argued by Keen and de Mooij (2012). Defining the scope of systemicness in a robust way is therefore indispensable for the design of appropriate measures towards such institutions.

Acharya et al. (2010) come with what I perceive as a measure of systemic importance, which is quite similar to the measure of externality proposed by Kocherlakota (2010), presented in subsection 9.3. According to their proposal, each financial institution should be required to purchase contingent capital insurance against systemic risk. The amount charged will allow the regulator to determine the relative systemic risk each firm poses to the financial sector, that is, how systemic it is. The authors justify the suitability of the cost of contingent capital insurance as an instrument for revealing which firms are systemic on their empirical results: using the proposed measure they find that most of the systemic risk is concentrated in just a few places⁸⁶.

⁸⁵ For example, Kocherlakota (2010) argues that the incentives of financial institutions to engage in risky investments will be unaffected by restrictions on their size.

⁸⁶ More specifically, their results show that over the sample period 5 firms were the source of 50% of the risk and 15 firms stood behind 92% of the risk.

16. Types of Proposed Taxes

Having looked at the various issues which, according to the literature, have contributed to the socially suboptimal allocation, the materialization of which was the recent crisis, and the potential role taxes and fees can have in addressing them, I now turn to concrete proposals of implementing the solutions, which have been put forward, in practice.

Whatever the name and goal of a tax or a fee proposed to be levied on the financial sector, all the considered variants fall into one of three broad categories: a Financial Transaction Tax (FTT), a Financial Activities Tax (FAT), and levies in one form or another on financial institutions' liabilities. The FTT is a unit tax levied on specific transactions. In contrast, the FAT is to be levied on the sum of profits and remuneration in the financial sector, which implies that it normally will be ad valorem, as will be the case with the liability levies. Examples of FTTs are all of the pre-crisis proposals mentioned in section 7., whereas implemented or contemplated taxation of remuneration in the financial sector are one application of FATs. The proposed stability fees are instance of the third category. Each type has its strengths and weaknesses. According to Keen (2012), FAT has attracted practitioners, while FTT has more political momentum in some places.

In the following sections I will present the general cases of FTTs and FATs and specific instances thereof: the scope for using levies for addressing the issues caused by the VAT exemption of the financial sector, and the countering of the negative effects of variable remuneration. Lastly I will look at the applicability of bank taxes and stability fees as means of ensuring sustainable risk and capital structures at financial institutions.

17. Financial Transaction Taxes

FTTs exist in many forms: securities transaction taxes (STTs), currency transaction taxes (CTTs, Tobin taxes), capital levies (a.k.a. registration taxes), bank transaction taxes, insurance premium taxes, real estate transaction taxes to name just a few. As this array suggests, they can, in general, be levied on any transaction involving one or more particular types of financial assets. The recent decades have been characterized by a process of gradual reduction and eventual abolition of many FTTs. Globalization has been the main driving force behind this development with the rationale that the increased mobility of capital across borders has meant that reduced capital costs are a necessary precondition for the competitiveness of a country's financial sector.

17.1. Getting the amounts of liquidity, transactions and speculation right

The financial crisis has spurred a discussion about the speculative motives behind the short-term trade in securities. It demonstrated convincingly that overtrading in the markets in the guise of too many transactions contributes to instability, inefficient capital allocation and weakening of the market's integrity. It has also been acknowledged that in practice ample liquidity and low transaction costs can provide market participants with incentives to decouple investments from fundamentals and economic reality, and instead focus on attempts to beat the market by guessing short-term price developments plus the increase in volatility. Thus, not all trading is Pareto-optimal from a social point of view. This superoptimality operates along two dimensions, at least: volume and frequency. As far as the former is concerned, Matheson (2011), for instance, argues that the fact of the matter has been that the volume of financial transactions has been in excess of what is needed to finance real activity (thus confirming the observation in *Business Week* mentioned in footnote 37).

The second dimension is perceptible in the tendency, a part of the general trend of shrinking horizons in the financial sector, for the duration of individual investments to fall steadily over time⁸⁷, which, thanks to the development of technology, has culminated in what is called high-frequency trading: an automated trading carried out by computer programmes trying to extract gains from arbitrage opportunities lasting a fraction of a second⁸⁸. As a result of the crisis, there has been recognition in the literature that if short-termed investors and fund managers are the dominant group affecting price formation, there is danger that the efficacy of the market will be weakened and that it will be easier for imbalances and price bubbles to form.

The superoptimality in trading is the core of the argument, echoing the ideas espoused by Tobin, of employing FTTs as a means of discouraging excessive speculation, especially combating hyper-frequent trading, targeting short-term round-trip flows. This reasoning faces two problems: one conceptual and one cost-related.

The former has to do with the ambiguity of various aspects and phenomena of the financial sector. More specifically, there are no clear-cut criteria about what useful and useless activity is, neither about the optimal level of liquidity. This translates into the lack of measures which specifically deal with useless activities, so that there is very real danger that attempts to curb them will affect negatively useful activities.

The second problem, highlighted already by Keynes in his *General Theory*, is the main trade-off involved whenever securities transactions are taxed: the benefit of curbing speculative activities comes at the cost of making real enterprise more difficult by reducing liquidity in the markets. Liquid markets (are supposed to) allow diversification of risk and facilitate the raising of capital, enabling the undertaking of large investment projects, and benefit savers through increased returns. At the same time, a liquid market can decouple investment from the fundamentals of an asset and then the near term returns will be based on short-term capital gains, opening the possibility for, and even inviting, speculation.

17.2. Alleged advantages

FTTs have been a controversial topic. Within the scope of the current composition, their potential role is as a tool for the regulation of financial markets. Their proponents claim that they can indeed be used to meet that goal – by reducing short-term speculation, FTTs will reduce the risk in financial markets and prevent the formation of asset price bubbles. This will further confer efficiency gains because resources will not be diverted to wasteful activities. By reducing profits and hence remuneration FTTs will cause the financial sector to contract, thereby offsetting (although not completely) the effects of the implicit subsidy it enjoys. Lastly, market volatility and asset mispricing will also be reduced. FTTs can further be used to achieve other goals, which are outside the scope of the current thesis, but which nevertheless matter when overall efficiency analysis is carried out, such as revenue-raising⁸⁹. A practical reason for support is the ease with which such taxes can be collected.

⁸⁷ For instance, NOU (2011:1, p.170) cites another paper by Andrew Haldane from 2010, “Patience and Finance”, according to which the average duration of the possession of a share of stock in the USA has fallen 12 times over a 67-year period: from 7 years in 1940 down to 7 months in 2007.

⁸⁸ NOU (2011:1) cites the same study by Haldane from 2010, according to which this type of trading stood for 70% of share trading in the USA and between 30% and 40% in Europe.

⁸⁹ Some of the controversy around Tobin’s tax was based on its misinterpretation as a revenue source for, among other things, the funding of globally beneficent activities. Tobin was dismissive of the idea and clear about that his proposal was to be evaluated on the basis of its potential use as an incentive-correcting tool.

17.3. Alleged disadvantages

The opponents of FTTs, on the other hand, see them as a rather blunt tool to curb the excesses of the financial markets (at the rare occasions when the former admit that the latter exist). Keen (2012) and Cottarelli (2010), for instance, argue that they are not suited to be used as tools of either financial sector regulation, or for crises resolution, with IMF (2010) rationalizing the former on the fact that FTTs are directed at none of the main sources of systemic risk, such as size and interconnectedness, and the latter on the volume of transactions' reflecting far too inadequately the benefits enjoyed by and the costs imposed by financial institutions operating in an environment of a resolution mechanism present. Shaviro et al. (2011) go even further by arguing that none of the claimed benefits in theory is substantiated empirically. Besides that, the opponents of FTTs claim that the latter will result in lower asset prices, lower returns to saving, and hence higher cost of capital for businesses, reduced liquidity, increased volatility, distorted financial markets and invitation for tax evasion on a large scale⁹⁰. Its incidence may well be mostly outside of the financial sector, thus making it poorly-suited to extract rents generated in the industry. In response to these objections, the advocates of FTTs have called for lower rates than what Keynes envisaged and Tobin proposed (the latter proposed as an example a tax of one percent in his 1978 article, against modern proposals going as low as half a basis point), in an attempt to mitigate the impairing of liquidity and movement to other jurisdictions.

In general, all taxes on transactions are routinely attacked on efficiency grounds: by discouraging mutually beneficial trades they reduce expected welfare. More specifically they disrupt the efficient pattern of ownership which requires that assets be owned by the people who value them the most as *The Mirrlees Review* (2011) argues. This line of reasoning obviously ignores the acknowledged fact that sometimes transactions are not a means to an end, but an end in themselves⁹¹.

17.4. Theoretical results on FTTs

Theoretical models confirm that increased transaction costs lower asset prices, the reason being that investors require higher return from the securities to keep them in their portfolio, when faced with higher cost of purchasing or selling one. The valuation premium on liquidity can be significant. Higher transaction costs raise the cost of capital for entities issuing taxed securities. A FTT has similar effect on the cost of capital as an increase in the rate of discount. In the model presented in Matheson (2011), securities values fall sharply the higher the tax rate and the shorter the holding period, while the impact on the cost of capital in his calibration drops quickly with holding periods' increasing, while the tax rate does not have much impact.

17.5. Empirical results on FTTs

According to the same reference, empirical studies of the impact of FTTs on financial markets mostly confirm that they reduce asset prices. Further, they confirm that there is positive correlation between the frequency of share trading and the impact of FTTs on a company's cost of capital. Higher transaction costs also lengthen the holding periods for securities. This has the effect of offsetting the impact of FTT on the cost of capital in equilibrium. With corporate bonds being traded normally less frequently than stocks the impact of FTT on corporate borrowing costs is likely to be smaller than the effect on stocks. The overall impact of a small FTT (up to 2 bps in Matheson (2011)) on corporate cost of capital would be modest.

⁹⁰ Citing all of the above reasons the committee behind the otherwise comprehensive NOU (2011:1) has not found it necessary to concentrate on FTTs, which is indicative of the rather strong opposition against them.

⁹¹ As an example consider the following quote: "...it [FTT] will fall arbitrarily heavily on those who, *for whatever reason*, engage in more transactions. [emphasis added]" (*The Mirrlees Review* 2011:152). Later in the report it is conceded that there are some cases where transactions need not be efficiency-enhancing, but the impression is that this is more of a minor nuisance, rather than a serious problem with the way the financial system operates.

17.6. Effect of FTTs on volume, liquidity and volatility

Theory predicts that FTTs reduce trade volume by making some transactions unprofitable. While, according to Matheson (2011), empirical studies find support for this conjecture for a broad range of elasticities across markets, this need not be the case: the impact of FTTs on trading can be negligible or absent, if there is regulation in place which limits the possibility of moving trade to other jurisdictions, or if the tax base is sufficiently broad. A potential problem which follows from reduction of trading is the very likely slowdown of the process of price discovery, because new information will then be incorporated into asset prices with a lag, thus actually increasing the possibility of an asset bubble's forming⁹². Matheson (2011) cites some empirical studies which seem to support this. Lastly, if the traded volume is reduced, liquidity should decline.

Empirical studies, however, do not provide any conclusive results of the effect of FTTs on liquidity: Matheson (2011) cites works on the topic which find that the impact can be positive or negative, depending on the market microstructure. Liquidity concerns can be avoided if a FTT based on realization of states of nature is employed. This comes at the price of greater distortions in behaviour.

According to Matheson (2011), FTTs can affect two types of volatility, which often are not clearly differentiated in the literature: price volatility in the short term and the boom-bust cycles gripping the market in the long run. They need not be correlated. While both types of volatility are taken into account by market participants, because they distort the price signals about fundamental asset values, it is the longer cycles which cause most concern from social point of view because of the far-reaching macroeconomic effects of bubbles and crashes they entail.

As he writes, in theoretical models there is an ambiguous relationship between FTTs and price volatility in the short term. As is the case with liquidity, it depends on the market microstructure. While it decreases liquidity and potentially volume of trading, the volatility in the presence of FTTs can increase or decrease. This inability of FTTs to discriminate between lower stabilizing (by arbitrageurs and informed traders) and destabilizing (by "noise traders") volatility is one of the main reasons for their rejection as a policy instrument.

The theoretical ambiguity, Matheson (2011) writes further, has found some confirmation in empirical studies: they find none or a positive effect of transaction costs on short-term volatility. There is, however, some evidence that trading activity itself generates short-term price volatility, so if transaction taxes depress trading activity, they could also neutralize it as a source of short-term fluctuations in prices. Further, *The Mirrlees Review* (2011) mentions that empirical research has not found any clear link between speculation and volatility, even though reduction of the latter is normally deemed necessary to make the former profitable, reflecting a proposition by Friedman that speculation must act as a stabilizing force on prices, lest it become unprofitable.

As far as the long-term volatility is concerned, "[t]here is currently a lack of research on the relationship between transaction costs and the market cycle." (Matheson 2011:15) Still, observations seem to point that FTTs would not prevent market swings from happening. This is because, according to the same reference, the boom-bust cycle is mainly explained in the literature with excesses of the credit cycle, that is, through the operation of the financial accelerator. Outside the mainstream, Minsky and the Theory of Long Waves point to volatility as a potential trigger of financial crises, even though it is not a cause for them by itself. So, as far as the more important long-term volatility is concerned, more research will be needed before any definitive conclusions can be drawn.

⁹² It is a completely different matter that historically there has not been a single instance of a bubble's being discovered until after it has burst, implying that the price discovery mechanism was not working properly anyway.

17.7. Incidence

The issue of incidence is important for two reasons. First, it is because of the distortions the substitution effect of each tax causes. Second, it is because one has to be sure that the residual payers of the tax are also the intended ones. According to *The Mirrlees Review* (2011), the incidence of a transaction tax is unclear, which, apparently, is a major problem with it.

According to Matheson (2011), the incidence of a FTT will be determined by the relative magnitude of two offsetting effects: reduced value of taxed securities and increased holding period. There is agreement in the literature that the burden of a FTT will not be confined solely to the financial sector, because it can be reasonably expected to find ways to pass the increased costs on to its customers, more so the wider adoption of the tax is, as this facilitates the transfer of outlays to customers; this is also the case with bank taxes and the stability fee to be discussed in later sections. It will, thereby, disproportionately burden sectors and activities heavily engaged in trading or issuing securities besides the financial sector itself, such as pension funds, public corporations, firms engaged in international commerce and public entities (unless transactions with government bonds were exempt from the FTT). It will therefore hit particularly hard the wealthier (who tend to sit on a lot of shares) and older taxpayers, because as savers they will have to face lower returns⁹³.

Further, Matheson (2011) notes, a FTT is like a capital tax: both drive a wedge between the cost of capital for the issuer and the net return on the security to the investor. The effects of the wedge will depend on the relative elasticities of the supply and demand for capital. In the textbook-cliché “small, open economy” with its perfectly elastic supply of capital, the imposition of a FTT will be the cause of capital outflow until its scarcity would raise its price up to the world market level, thereby burdening the owners of immobile factors, such as land and labour, in the form of reduced productivity. Thus, the cost of capital to firms will increase both directly — through higher borrowing and factor outlays — and indirectly — with the unclear effect on short-term volatility uncertainty will increase — so that the net effect may be decreased social welfare.

The latter can also be the result of the very real possibility of FTTs’ having perverse, counter-intended results. For instance, the increased uncertainty they will cause may either induce more conservative behaviour, which is the intended goal, but might also encourage misbehaviour, which is what one wants to avoid. Closely related are the claims in certain studies that the greatest share of the tax might end up being levied on long-run orientated, risk-averse investors, while those whose behaviour the tax is supposed to correct might be able to pass the extra burden away and actually engage in riskier activities by speculating in exempt or otherwise unregulated securities. On the other hand, Stiglitz (1989) argues that there will be no significant effect on long-term investors, because of a turnover tax’s, which he considers and which according to Cottarelli (2010) is similar to a FTT, “automatically phasing itself out for long-term investments”. (Stiglitz 1989:112)

Beyond all that, the committee behind NOU (2011:1) points out that the introduction of a global fee levied on currency transactions (which has been one of the most discussed transaction fee forms) will have negative consequences for Norway and other small currency areas, because in such small, open economies there are many who trade with foreign countries and are exposed to currency risk. The need for currency exchange in such countries will, therefore, be larger, than is the case in the USA or in the Eurozone.

17.8. Distortions

According to Matheson (2011), in general, transaction taxes are considered more distortive than other taxes levied to achieve a particular goal. He then goes through numerous undesired effects of

⁹³ As this already is the case because of the low-rate policies of central banks, potential introduction of FTTs will further reduce welfare for these two groups.

FTTs that have been pointed out in the literature. First, like a capital income tax, they can either reduce or increase current consumption, because they will distort the production process by raising the costs of some forms of capital. This has to do with the fact that finance is an intermediate good and as such should not be taxed. Second, a FTT taxes gross flows, so that unless the tax is creditable to those using financial services as intermediary products, it would have a “cascading effect” (Matheson 2011), leading to multiple taxation on some transactions (as IMF (2010:18) puts it: “tax being charged on values that reflect the payment of tax at earlier stages”). Third, connected with the previous one is the problem that the same product (understood as an allocation of assets and risks among investors) can be produced using different combinations of transactions, each of which commands different amounts of the transaction tax. This means that it is not possible, in principle, to design a FTT in such a way that it imposes the same tax burden on all types of financial transactions which deliver the same economic outcome: some paths to a particular outcome will be subject to multiple taxation, while others may be subject to none at all. This is likely to hit particularly hard assets which have long maturities or are frequently traded. IMF (2010:19) refers to this as financial transactions taxes’ being vulnerable to “avoidance by engineering”. Also connected with cascading are the incentives to avoid the tax through in-sourcing, potentially leading to financial institutions’ becoming bigger. A further issue with gross taxation is that it does not provide for proportionality between the tax burden and the surplus realized in a transaction: even loss-making transactions will produce tax liability. It is, however, this imposition of positive costs in good *and* bad realizations which makes FTTs suitable as tools for discouraging risk-taking – a similar point can be found in Stiglitz (1989:112), who argues that a tax with no provision for loss deductibility precludes arbitrage opportunities which lead to losers’ being subsidized and winners’ being taxed. Together with the electivity of the tax — due if one voluntarily engages in trading — this distorts trading relative to an (ideal) income tax on capital gains and a wealth tax (which is non-elective).

17.9. Practical issues

There are various issues when it comes to the implementation of FTTs. First, they are easily levied, because the taxed transactions require that the parties register their ownership of the items in question; *The Mirrlees Review* (2011) cites the UK experience with “stamp duties” as a relevant example. According to Matheson (2011), FTTs have advantages over other taxes in terms of administrability: they may be cheaper to collect, but the cost will increase significantly if levied on over-the-counter transactions as well.

The theoretical feasibility and (conditional) advantages of such taxes would suggest that most of the attention would be accorded to their particular design. Two aspects are central in this regard: the tax base and the tax rate.

A broader tax base will reduce the possibility for avoidance and the distortion of financing and investment decisions. To attain the former, the relationship between taxed and untaxed instruments must be considered with a view to including as many potential substitutes for taxed securities as possible, while the latter would require levying the tax on both debt and equity instruments. Thus, to achieve the broadest possible base, FTTs should ideally apply to all types of traded securities: equity, debt, foreign exchange and their derivatives; as Matheson (2011) writes, the latter because if left untaxed, trade would migrate from spot to derivatives markets, increasing both risk and leverage and thus doing exactly the opposite of what the tax is supposed to do. Various important aspects will have to be taken into account when it comes to determining the tax base for the last ones, such as the effect on leverage and opportunities for arbitrage. The broadness of the base will depend on whether both exchange-traded and over-the-counter instruments are taxed. While the inclusion of the latter will increase compliance costs, exclusion would provide incentives to trade more securities over the counter, which reduces transparency, as Matheson (2011) argues. In addition, while not mentioned in the literature, but, in my opinion, nevertheless important, such a “migration” is problematic, because over-the-counter markets are normally less liquid than exchanges. One way to

reduce the extent to which standardized products are traded in over-the-counter markets would be to impose higher rates on transactions conducted there.

A further important issue in the determination of the base is territoriality. It will have implications for the potential to evade the tax and its administrability; in fact, there is agreement in the literature that the effectiveness of the introduction of the tax will depend heavily on international cooperation and coordination – global or even multilateral imposition of FTTs will not jeopardize the stability of the financial system, because it will only entail income effects, thus reducing the scope for distortions. While the imposition of FTTs on transactions occurring in a given location is likely to drive trade activity to other jurisdictions, this need not be the case. Nevertheless, the increased cross-border integration of financial markets still presents a major challenge to unilateral introduction of FTTs. Financial activity entails agglomeration effects, so that established financial centres seem to face a less elastic tax base than other countries; international cooperation will have the effect of reducing the elasticity of the tax base for the latter ones as well, as Matheson (2011) points out.

As far as the tax rates are concerned, their determination will depend on what one tries to achieve. Whatever the goals one may pursue, one will have to consider whether to charge flat⁹⁴ or different rates in different markets depending on the elasticity of their base or their level of non-tax transaction costs. However, within the scope of the current composition — where dealing with externalities is the sole goal of such a tax — Matheson (2011) argues that derivatives may have to be charged heavily with respect of their low transaction costs, as leverage causes systemic risk⁹⁵. As he argues further, although transaction costs are generally higher in over-the-counter markets, there may still be levied higher rates on trades conducted at those venues as a means of pushing trading to the more transparent and more liquid exchange markets.

Last, but not least, the eventual introduction of FTTs is a legal matter, to the extent that they may be viewed as obstacles to the free movement of capital, wherever it is imposed by law or some binding agreements.

17.10. Examples of implementation

Various types of FTTs have been/are levied in different countries, however none of them for the explicit purpose of reducing speculative activities, which, of course, is easy to explain by the fact that their use is situated in the pre-2008 period, when the dominant mindset was different.

A widely-cited example is the Swedish application of a financial transaction tax, popularly known as *Valpskatt*, in place between 1984 and 1991. The experience was a failure, which is explained by the poor design of the tax. It is an empirical confirmation of the postulated negative effects of a unilateral imposition of FTTs listed in, for instance, NOU (2011:1).

A post-crisis proposal, which also among the other goals aims at complementing the existing regulatory framework to avoid future crises, is the EU's FTT. It is supposed to levy 0.1% of exchanges on shares and bonds on secondary markets and 0.01% on derivatives contracts, in effect taxing all

⁹⁴ The literature does not accord attention to the question whether a flat tax rate will affect incentives to engage in certain trades. Slutsky's theorem postulates that the total effect of a price change is the sum of a substitution effect, which is always negative, and an income effect, which can go either way. It is plausible that when there are high amounts at stake, the income effect may dominate, rendering the tax ineffective exactly in the cases it is intended to protect against.

⁹⁵ As far as derivatives are concerned, there is a second dimension to take into account, besides frequency and volume of trading, namely total exposure to a particular type or particular originator(s) – indeed, this is what we have read about in the news: that a given financial institution is at risk, because it has exposure of a(n) (un)certain amount to some troubled derivative contracts, not because it has traded far too often. This second dimension is not something FTTs can be used to address. Instead, the issue can be dealt with through FATs.

transactions between financial institutions involving at least one party resident in the EU, thereby making it evasion-proof. Its targets include automated high-frequency trading and excessively leveraged derivatives. It will not affect the real economy directly, since home mortgages, loans to small and medium enterprises, insurance premia, currency exchange on the spot market and securities issues in the primary markets are exempt. However, according to an official study by the European Commission, the cost of capital is predicted to increase, thus indirectly affecting non-financial economic actors adversely. In addition there is an estimated reduction of the EU's GDP in the long run by 0.5% to 1.76%, depending on whether a good or a bad scenario is realized.

17.11. Judgement

The opinion expressed in Matheson (2011) is that FTTs are an inefficient instrument for regulating financial markets, because they do not offer as specific a remedy as other taxes and regulatory devices. More specifically, they are ineffective in preventing the formation of bubbles, because the consensus transcending both mainstream and non-mainstream literature attributes the latter to excessive leverage, while FTTs primarily affect short-term trading. The acquisition of excessive leverage can be discouraged through other means, such as increased collateral and margin requirements, or a tax on transactional leverage. The treatment in Matheson (2011) does not consider the political and administrative issues with the implementation of FTTs; taking them into account would likely further strengthen the case against their use. This line of reasoning is continued in Shaviro et al. (2011) where it is argued that FTTs are not appropriate to either raise revenue in an efficient way (compared to other ways of doing so), or to contribute to more equitable outcomes (however they may be defined). While these two uses are of no direct relation to the current writing, the argument in the paper shows that welfare will most likely be reduced if FTTs are introduced⁹⁶, without aligning the behaviour of the financial market actors with the socially optimal. The authors behind *The Mirrlees Review* (2011) also doubt that transaction taxes on the sale or purchase of securities will curb excessive speculative activity and for this and distortion reasons argue against their use. According to NOU (2011:1), the combination of lack of agreement in the scientific community about whether the imposition of fees on financial transactions is suited to promote financial stability and counter undesired activity in the financial markets, the negative effects such levies have and the difficulties in imposing them in an effective way does not make FTTs a viable option. Keen (2012) also gives the impression of the dominance of uncertainty with respect to claimed benefits, feared costs and implementation issues.

FTTs are not quite useless, however. The recent crisis revealed that many institutions relied too heavily on short-term funding. FTTs can be employed to discourage them from doing so, especially in the case of non-banks, because they are not subject to capital requirements.

Further, Matheson (2011) cites the view of the European Parliament from 2010 that a low FTT can be used as a provisional tool to substitute for the inadequacy of financial regulation in the short term. The core of the argument is that the quick pace of financial innovation and trading create risks which are often poorly understood by both market participants and regulators (a point stressed several times in NOU (2011:1)). A FTT can be used to slow the pace down. The tax can be used provisionally until other taxes or regulatory measures, better aimed at dealing with the risks the respective practices entail, are established. However, a limitation to the effectiveness of this proposal is the relative illiquidity of new financial products, because it is likely to prevent a FTT from having any

⁹⁶ For example, the average realized tax revenue from the Swedish FTT was 30 times less than what had been expected. Further, the reduced trading volumes (as a result of the poor design, which imposed the tax only on domestic brokerage services, close to 60% of Swedish trading had moved to London shortly before the complete abolition of the levy) diminished the base which capital gains taxes were levied on, in effect nullifying any income the tax produced. Lastly, the tax increased the cost of government borrowing, because fixed-income securities were not exempt.

significant impact on the pace of financial innovation, although it could still reduce trading abuses hidden in high-frequency trading.

18. Repealing the VAT Exemption of the Financial Sector

Financial services are exempt from VAT in many countries. This non-neutral and preferential treatment raises many issues⁹⁷. The one among them relevant for the current composition is the effect this arrangement has on the financial sector's size and volume of supply it offers: being equivalent to a subsidy, the VAT exemption implies underpricing and hence oversupply of financial services; the additional rents this creates gives incentives to the sector to increase in size.

The effect of VAT exemption in pecuniary terms is significant. As an example, consider the numbers for Norway: NOU (2011:1) cites the IMF according to which the value added by the financial sector in Norway is approximately 2.7% of GDP (based on continental GDP at the end of 2009), if profits and remunerations disbursed are used as a proxy for measuring it. Using the standard VAT rate of 25%, this means that the exemption leads to a tax revenue loss of some 12-13 milliard NOK. NOU (2011:1) cautions that this number is highly speculative, since it is neither very certain that the measure used gives a realistic picture of the value added by the financial sector, nor does it take into account the deduction of VAT on purchases of inputs, which would be applicable if the financial sector were subject to the tax. That notwithstanding, for the purposes of the current composition this figure is more important as a very rough measure of the fraction of the part of the implicit subsidy to the domestic financial sector due to undertaxation. As was mentioned in footnote 84, the report estimates the value of the implicit state guarantee to the Norwegian financial sector to be around 5 milliard NOK, bringing the total amount of the implicit subsidy to the domestic financial industry to 17-18 milliard NOK. With close to 2/3 of this sum being due to VAT exemption (which is not the full measure of the financial sector's undertaxation), it can be seen that, at least in pecuniary terms, improvement of the taxation of the financial industry will achieve more in reducing the implicit subsidy it enjoys.

18.1. Reasons for exemption

Various reasons for the existing arrangement have been put forward, besides the most obvious one, namely lobbying. The top spot is occupied by the lack of a sale, on which VAT could be charged: it is claimed that under the predominant consumption-type invoice-credit form of VAT there are technical difficulties to define the tax base, because there is no explicit price for the provision of loans or the acceptance of deposits. In the aftermath of the crisis some ways around this problem have been proposed: for instance, *The Mirrlees Review* (2011) argues that since the tax authorities normally use an invoice with a reported price to charge VAT, one starting point to determine the tax base is to compare the interest rate spread (between loans and deposits) to a benchmark interest rate to determine the value added in the financial sector; the concepts of cash-flow taxation and tax calculation accounts are based on this approach.

The prime candidate for the next champion of the defenders of the exemption is the production efficiency theorem of Diamond and Mirrlees (1971), which postulates that goods and services used as intermediate goods in the production of other goods and services should not be taxed.

⁹⁷ While the discussion on the topic is not new — NOU (2011:1) cites an earlier NOU (1990(!):11) which had come to the conclusion that “the significant values created by finance and insurance in principle appear well-suited to be brought under the authority of the VAT law” (A rather literal translation of the Norwegian text: “...de betydelige verdiene som skapes gjennom finansiering og forsikring i utgangspunktet synes godt egnet for å trekkes inn under merverdiavgiftlovens område...” (NOU (2011:1, p.175))) — an extra dimension has been added to it by the recent crisis, namely that the repeal of the exemption can now be regarded as a tool to advance and promote financial stability.

The next reason is the impossibility for banks to reclaim VAT on purchases from registered traders, the latter being necessary for the provision of the financial services the former offer, as argued in *The Mirrlees Review* (2011). Next come the high mobility characterizing the financial sector and the freedom of capital movements in our days with the rationale that differences in taxation practices generate a tendency where providers of financial services tend to gather in jurisdictions with liberal regulations⁹⁸. Further, if it is difficult to identify which inputs can be traced back to exempt activities, there is the incentive to use combination of taxable and exempt activities. Lastly, as cost minimizing entities, financial services providers will tend to lean to zero-rated inputs and vertical integration, thus avoiding VAT payments on purchased inputs as much as possible.

18.2. Distortions caused by the VAT exemption

Several distortions which the VAT exemption enjoyed by the financial sector causes have received attention in the literature. The arrangement produces an allocation in which households and businesses are “distorted” through distorted relative prices away from the more expensive goods and services subject to the VAT and towards the financial sector, while for reasons of allocative efficiency, resources used in the production of financial services should in theory be charged VAT in the same way as resources used in other sectors are, even though in practice the input-output structure of the economy will have a major bearing on the importance of allocative effects. The exemption provides a tax advantage to private consumption of financial services, and a tax disadvantage to uses in production, so that the former are undertaxed, while the latter are overtaxed, thereby breaching the neutrality principle on which the fee structure is ideally founded. In this way the exemption affects both consumption and the business structure, leading to a socially suboptimal allocation, in which too many resources are directed towards the financial sector. Further, there is agreement in the literature that the VAT exemption favours in-sourcing of production, because it is less costly for the financial sector to itself produce the inputs it uses. As a result, it becomes bigger than necessary. In addition, according to *The Mirrlees Review* (2011), the degree of financial sector undertaxation due to its VAT exemption grows as the rate of the tax is raised, so that if, as is likely to be the case, there is a positive feedback between the size of the implicit subsidy and the size of the financial sector, VAT increases elsewhere (as have either been contemplated, or already implemented as way of dealing with problematic government finances) may potentially make the sector even bigger by increasing its subsidy, leading to an even more distorted allocation. Lastly, in an international context, VAT differences across countries can cause distortions where services will be moved to areas with lower, or no tax at all.

18.3. Difficulties with and potential drawbacks of repeal

The major difficulty is to determine what to tax; indeed, this is part of the larger problem of precisely defining what exactly constitutes financial services. The challenge of defining the tax base has much to do with the currently far too inadequate ability to measure the value added created by the financial sector – while it is possible to compute VAT for explicitly priced services, it is extremely difficult to do so for each transaction involving implicit price components, based on various margins.

Cross-border cooperation and coordination are important, because in an international context the place where taxation takes place is essential – currently VAT is levied on destination basis that is, paid by the buyers of goods and services according to VAT system in the country of purchase⁹⁹. The present arrangement has the benefit of making tax-evasion of services offered locally difficult, but creates a number of inefficiencies with cross-border purchases and problems of administration. In the absence of effective restrictions on capital movements, international coordination will be

⁹⁸ According to NOU (2011:1, p.176), the consideration for the tax and fee framework for financial services in comparable countries was decisive for the Storvik-committee’s — the one behind NOU (1990:11) — recommendation to exclude financial services from the VAT system for the time being.

⁹⁹ The alternative is to implement VAT on origin basis, where the tax is levied in the country of production.

necessary to keep the risk of reallocation in reaction to the tax low. Still, according to NOU (2011:1), considerations about the competitiveness of a country's financial sector are in principle no obstacle should it decide to impose a VAT, even if other countries do not. For this to work, however, one will have to find effective ways of taxing the import of financial services to local residents. Even if such techniques are employed, the turnover of the more mobile services will still move to countries with lower or no VAT. Lastly, significant administrative and judicial issues will have to be dealt with, for instance, the VAT exemption of financial services is mandated by EU rules, as mentioned by *The Mirrlees Review* (2011).

Once a reasonable tax base is determined one will move on to consider the tax rate – as far as VAT is concerned this revolves around the question whether it should be uniform or differentiated. The authors behind *The Mirrlees Review* (2011) are rather sceptical about the use of differentiated rates of indirect taxes, even though, in general, they admit that efficiency reasons (captured by the inverse elasticity rule, or its more general version the Ramsey rule) would support the case for such structure of the levies, because a differentiated rate system has the benefit of providing the authorities with the opportunity to be as close to the efficient tax rates over time as possible, but at the cost of being unwieldy. They point out two benefits of a uniform tax rate: first, it avoids complexities, lobbying and short-term political considerations which are always present with differentiated rates; second, administration and compliance will be easier. On the other hand, correcting for externalities is an argument in favour of a differentiated VAT, because it should reflect the different damage and/or benefit it causes.

The imposition of VAT on the financial sector, once the issues surrounding the implementation have been resolved, will involve certain drawbacks, at least in the short run. Several such challenges are cautioned about in *The Mirrlees Review* (2011). First, the volume and complexity of financial transactions in modern economies may prove a serious challenge to tax authorities, pushing them to, and even beyond their limits of administration. Second, the time necessary to devise and implement such a fundamental change makes attempts to avoid the tax before it enters into force a very real possibility¹⁰⁰, because while it is not impossible, it is highly improbable that such process would be kept secret; in addition, this also opens up the possibility of spreading the news only to select few insiders. Likewise, the effectuation of a change with such profound effects would possibly take a long time to get used to. Lastly, any levy on purchases by banks from registered traders will be successfully passed on to the financial intermediaries' customers in the form of less favourable interest rates.

Within the scope of improving the regulation of the financial sector, Matheson (2011) cautions that even if VAT is extended to cover all fee-based financial services, this will only partly rectify the distortions caused by the exemption, because the financial industry will still be undertaxed. Therefore, the subsidy to the financial sector, and hence the unwanted effects it entails, will not be completely eliminated.

18.4. Concrete proposals for rectification

Despite the aforementioned yet-to-be-resolved issues, concrete proposals to impose VAT on financial services have been put forward. One particular example, discussed in among others *The Mirrlees Review* (2011) and Keen, Krellove and Norregaard (2010), is what is referred to as “cash-flow VAT”¹⁰¹. The proposal is to treat certain cash inflows to a bank — deposits, interest on loans, repayment of loan principals — as sales, on which VAT is due, and certain cash outflows from a bank

¹⁰⁰ For instance, news of the Swedish FTT was leaked prior to its effectuation, affecting markets adversely already under the pre-existing regime.

¹⁰¹ Other examples are mentioned in Keen (2011).

— loans made¹⁰², interest on deposits, withdrawals of principal — as input purchases, on which VAT can be reclaimed. Among the cash flows which will not be subject to the proposed tax are equity transactions, because they are “final goods” produced by suppliers of financial services, but, depending on the goals pursued, should be subject to either a capital gains tax (if one aims at addressing the undertaxation of the financial sector), or to a kind of a Tobin tax (if the purpose is to counter speculation in equities), or in the case of dividend payments, which do not qualify for a FTT, to a FAT. Further, the proposed VAT would obey the principle that intermediate goods should not be taxed, implying that it will be levied only on cash flows to and from consumers, while provision of financial services of the business-to-business type will be exempt. The base of this proposed VAT would be broadened to include more sophisticated financial products, so that the possibility of using long, complex legal definitions of what can be considered exempt is eliminated. The authors behind *The Mirrlees Review* (2011) importantly emphasize that this extension of the tax base does not imply that the tax rates will be low — a proposition easily understood with a view to curbing the excesses and whims of the financial markets, where the goal is a changed allocation as a result of the substitution effect, thus contrasting with the conventional principle of broad-base, low-rate taxes, which reflects the idea of minimizing the impact of the substitution effect, when the goal is to raise revenue. As the report reads, the proposed cash-flow VAT can be realized on transaction-by-transaction basis (i.e. as a FTT) or by examining the consolidated account of financial entities for the period of interest (i.e. as a FAT). If the second option is chosen, there is the possibility of calculating the tax base as the sum of profits and wages, which is just another way of calculating how much each entity is worth, (rather crudely) reminiscent of the way GDP is defined in national accounting. Keen et al. (2010) mention two weaknesses of the approach: it may be the source of cash flow problems for businesses because of principal amounts’ being taxable, and it may be difficult to administer in the case of large-volume cash flows. Workable fixes to these problems — called “Tax Calculation Account” and a refinement of it, called the “zero-rating approach” — have been put forward; see Box 1. in Keen et al. (2010:121) for a detailed account.

The proposed cash-flow VAT, or some of its equivalents, will eliminate or, taking heed of the warning by Matheson (2011), offset to a certain extent the current distortion whereby financial services to households are undertaxed, while those to businesses are overtaxed, thereby addressing the issue of superoptimal use of resources in the financial sector. Within the scope of the current composition, the potential benefit this may confer consists in the reduction of lending to potentially bad risks, caused by the decreased after-tax rent from interest on loans accruing to the lenders. The cost comes in the form of a possibility of credit rationing arising in equilibrium, which is welfare reducing, as some potential good risks will be denied credit. The net result will be ambiguous: it may well happen that in an attempt to obtain the same profit, banks simply compensate the fall in interest income by extending more loans. While this alleviates the credit rationing problem, it may have the counter-intended effect of increasing the number of risky loans.

The proposal has received some empirical support: *The Mirrlees Review* (2011) cites the results from experimental pilot studies of volunteer banks in Europe in the 1990s, where the cash-flow VAT has proved to be conceptually robust and administratively manageable, but its high compliance costs make it currently impracticable, according to Keen et al. (2010).

As far as Norway is concerned, while the committee behind NOU (2011:1) is of the opinion that it is unfortunate that an entire sector is exempt from VAT, and the reconsideration of this state of matters occasioned by the financial crisis notwithstanding, it expresses the view that it is likely going to be very difficult to find a suitable model for imposing VAT on the domestic financial sector. A system tailored to the needs of the country will likely require large development outlays and its

¹⁰² However, *The Mirrlees Review* (2011) warns that allowing VAT reclaim on loans made is problematic, because it will have the same effect as exemption.

implementation will be difficult without significant adjustment and social costs. For these reasons, the report suggests that one focus on alternative solutions and international cooperation.

18.5. Empirical results

According to Buettner and Erbe (2012), there has been only token discussion on the empirical estimation of the revenue and welfare effects which the repeal of the financial sector's VAT exemption would produce, and whatever is extant deals primarily with revenue effects. The findings of their article, which is a contribution towards the urgent necessity of filling the void, are that both revenue and welfare effects are much less promising than indicated by earlier research. Based on general equilibrium approach their study shows that there will be both direct effects on tax revenue and indirect effects through changes in consumer and producer prices of repealing the financial sector's VAT exemption; for instance, one component of the effect on revenue is the potential changes in employment resulting from a change in labour supply, the elasticity of which will determine the impact on the labour market of the effects on the overall price level. Further, the welfare effect of the changes in the demand for intermediate inputs will depend on the cross-price elasticities: if there are none, then there is a tiny welfare loss, if they are positive, there is a small welfare gain.

The paper by Buettner and Erbe (2012) contains an estimation of how the effects of repealing the VAT exemption of the financial sector will depend on how much of the tax the industry is able to pass on to its customers. Based on German data, their calculations show that if nothing can be passed on there will be a 0.7% loss of total VAT revenues in the base year. In the other extreme, if all of the tax can be passed on there will be a 2.7% increase in total VAT revenues in the base year. In between these two is what they regard as the most realistic one: the authors' estimation show that 58.8% of the tax will be passed on and the increase in total VAT revenue in the base year will be 1.2%.

Within the scope of the current composition, the relevance of the results in Buettner and Erbe (2012) stems from their showing that it is the customers of the financial sector who will have to shoulder most of the bill for the elimination of its implicit subsidy, should the industry's VAT exemption be repealed. More generally, the introduction of VAT on the financial sector will have mostly welfare-reducing effects¹⁰³. So, in conclusion, Buettner and Erbe (2012) argue that, all else equal, the technical difficulties with the implementation of VAT on financial services taken in consideration, combined with uncertain effects on welfare do not seem to warrant the repeal of the exemption for the financial sector. This conclusion, however, follows from the setup of their paper, which does not focus on the externalities the financial sector imposes on the rest of the economy. It would be an interesting extension to their study to take the latter in consideration and see if this affects their conclusion.

18.6. Alternatives

The two other reviewed works which explore thoroughly the case for imposing VAT on financial services — NOU (2011:1) and *The Mirrlees Review* (2011) — concur in their conclusion that there are too many difficulties and uncertainties with the direct implementation of such a policy. The authors behind both studies believe that a more viable option for correcting for the VAT exemption of the financial sector is to use a variant of a FAT instead. Regardless of which approach should be chosen, it

¹⁰³ More specifically, the model in the study predicts a welfare loss from the deadweight loss induced by the introduction of the tax on the consumer prices in the financial sector; the empirical results, which throughout the paper depend critically on the amount of unrecoverable input taxes, confirm this. The welfare effect is that repealing the exemption will increase excess burden of taxation, whether the financial sector can recover input taxes or not. There is an additional distortion of labour supply, because the increased consumer price level outweighs the beneficial effects for the consumers through lower producer prices. Buettner and Erbe (2012) find a small welfare loss when the financial sector is able to recover 58.8% of the input taxes, this result being based on the assumption that the revenue changes are transferred to households in a lump sum fashion.

will be important to consider the interaction of the VAT with other taxes and fees imposed on the financial sector.

19. Financial Activities Taxes

FATs are the second major type of taxes which have received considerable attention in the policy debate in the aftermath of the recent crisis. In general, a FAT is a tax on the sum of profits and remuneration of financial institutions. The latter in turn is value added; hence the analogy with the definition of GDP mentioned before. The precise definition of profits and remuneration in each particular instance will determine the economic impact and potential revenue of a concrete realization of a FAT. Contrary to FTTs, FATs have been looked upon favourably; for instance, the use of FATs over FTTs is advocated in IMF (2010). In the reviewed literature there seems to be agreement that FATs are suitable instruments to employ so that the various goals policy makers may have towards the financial sector are attained: the ones relevant for the current composition, such as effective regulation and supervision, induced prudence, “fair contribution” by the sector to recover the costs for its bailout, to the extent this helps equalize private with social marginal costs; and others, such as equity and revenue-raising¹⁰⁴. Specifically, as far as the former set of objectives is concerned, FATs can be used as a tool to offset the second element of the implicit subsidy to the financial sector: the “implicit state guarantee”. Closely related is their potential application to deal with the undertaxation of the financial industry. Lastly, pecuniary measures to deal with the unfortunate incentives provided by the remuneration arrangements in the financial sector will involve some kind of a FAT.

19.1. *Proposed variants*

Shaviro et al. (2011) point to efficiency reasons favouring FATs, namely they will fall on rents. Indeed the rents targeted by all the proposed variants are the sum of profits in excess of normal return and remunerations above a certain threshold. The rationale is that both serve as proxies for excessive risk-taking and of a bias towards short-termism. In addition, as pointed out by IMF (2010), FATs will in effect tax net transactions of financial institutions, unlike FTTs, which will be levied on gross transactions. Whatever the particular form a concrete realization of the tax may take, it is important that the latter be activity-based, not institution-based; otherwise it may not achieve its regulatory goals¹⁰⁵.

As was mentioned before, each concrete realization of a FAT is suited to serve certain purposes. Three ways of implementing a FAT have received attention in the literature, each directed at a specific goal. Keen et al. (2010) refer to them as FAT1, FAT2 and FAT3 respectively.

The purpose of FAT1 is to address the undertaxation of the financial sector, primarily due to the inadequacies of the existing VAT system. As argued in Keen et al. (2010), Keen (2011) and *The Mirrlees Review* (2011) the net VAT due is by accounting identity equal to the sum of profits and all wages, and the latter can be used as the base of FAT1¹⁰⁶. The advantage of FAT1 is that will effectively tax all of the value added. While the effect of FAT1 on wages is straightforward, it should be neutral with respect to profits, in the sense that it does not affect investment decisions and

¹⁰⁴ There is scope for this, for instance, Keen (2011) expresses the belief that the revenue potential of such measures can be large. Matheson (2011) goes further by arguing that FATs will produce less distortion than FTTs in the process of collecting a given amount of revenue, because they tax net value added, not the gross value of transactions.

¹⁰⁵ Shaviro et al. (2011) cite as an example the potential exemption of the financial divisions of non-financial firms, which would encourage migration to them and thus distort the structure of the actual financial sector.

¹⁰⁶ According to Keen et al. (2010), the practice is referred to as “the addition method” for calculating VAT liability, contrasting the standard invoice-credit method.

marginal financing. Keen et al. (2010) mention three alternative measures of profits, under which FAT1 will have this neutrality property: two cash-flow approaches — the so-called R+F base (in which both real and financial transactions are taxed and deducted), and the S base — and allowance for corporate equity, with the latter's having the advantages of requiring little adjustment due to its closeness to the way businesses are currently taxed, and of retaining its neutrality even if the tax rate is expected to vary over time¹⁰⁷. Further, the same authors argue that VAT paid on purchases used in the provision of VAT-exempted services should not be creditable against the FAT1 liability. Lastly, they argue that the threshold level of activity above which an entity is subject to FAT1 should be set sufficiently low, so that no institution engaged in significant financial activities finds itself exempt. As a major theoretical drawback of FAT1, however, *The Mirrlees Review* (2011) and Keen (2011) point out the incidence of the tax's being undetermined. Another one, pointed out by Keen et al. (2010) is the risk of FAT1's cascading and causing production inefficiencies if non-financial firms' VAT liability is determined by the invoice-credit method. The same authors argue further that a FAT1 levied on a destination basis will exacerbate what they call "import bias" — the propensity to sell/buy financial services abroad rather than domestically — and lead to loss of revenue; as a fix to the problem they suggest levying the tax on the origin basis.

FAT2 and FAT3, on the other hand, are to be levied on the economic rents arising in the financial sector. As elucidated by Keen et al. (2010:130), these "inframarginal returns", as they call them, are part of the returns to capital (labour), together with the "normal return" — "risk free, pure time value of money return" (the wage a financial sector employee with given characteristics can earn in other sectors) — and risky returns (which, according to the three authors, may involve substantial discrepancy between actual and expected returns).

FAT2 is intended to tax away some of the rents to capital and labour accruing to actors in the financial sector. The amount raised can be then used to deal with costs of future trouble in the financial sector, in effect helping to close the gap between private and social marginal costs of the industry's activity. Keen et al. (2010) cite empirical evidence of these sector-specific rents' being significant, with too-big-to-fail-ness and informational asymmetries being among the primary reasons for these inframarginal returns' being so marked. Among the issues FAT2 would ideally help address, besides being an efficient way to raise revenue, is improperly designed remuneration schemes, which are blamed for exacerbating the moral hazard problem (in the form of increased risk-taking and short-termism) to which management is subject to begin with due to the separation of ownership and control. It is envisaged that FAT2 would be levied on the sum of profits and "high" wages. While the identification and taxation of economic rents to capital can be done using the same measures as with FAT1, the rent element of remuneration, more specifically an individual's human capital according to Keen et al. (2010), may not be clearly distinguished, which makes taxing it difficult in practice. To the extent that labour markets deviate from their perfectly competitive ideal (a source of uninterrupted lament for the free-marketeers), the rents to wages can be measured by the difference in payment for jobs similar or requiring the same skills elsewhere in the economy, or, more generally, to some benchmark based on figures from the entire economy. As far as the financial sector is concerned, Keen et al. (2010:134) argue that an appropriate proxy for wages in excess of the perfectly competitive ones would "include all payments ["wages, bonus payments, gains on stock options and other payment and gains"] accruing to employees", because any of the listed forms may enable them to appropriate part of the surplus income and rents generated by the industry. Further, theory suggests that some targeting of specific individuals may be necessary. Notwithstanding the fact that FAT2 can be designed in such a way that its incidence is on the rents to owners and managers of financial enterprises, its restrictive conditionality in application is a serious drawback. While efficiency reasons would suggest that FAT2 should be applied to all financial institutions, Keen et al. (2010) point out that administrative and compliance costs warrant subjecting to it only entities

¹⁰⁷ See 17.3.1. and 17.3.2. in *The Mirrlees Review* (2011) for an account and further reference on these.

which can reasonably be expected to be the source of significant rents, such as (potential) too-big-to-fails.

FAT3 is considered as an instrument for discouraging excessive risk-taking – Keen et al. (2010) explicitly stress its nature as a Pigouvian tax. Its tax base should ideally consist of high wages and returns to equity at a rate in significant excess of what is considered normal, with the idea of introducing sufficient amount of progressivity, such that excessive risk-taking is discouraged; more specifically, by reducing the net return of good realizations, the tax would correct the undervaluation of bad realizations stemming from the expectations that the associated costs will be passed on to others, which fosters risk-taking in the first place. In defining the base, one could choose to tax excess return to assets instead of equity, which could be justified on the ease for investors to move between debt and equity. On the other hand, taxing only returns to equity will contribute to reduction of the debt bias in financing, due to the favourable treatment of debt for corporate tax purposes. Keen et al. (2010) stress that for the incentives for risk-taking to be affected, losses and gains must be treated asymmetrically, for instance by eliminating the opportunity to write off losses against other taxed income – I have made a similar point towards the end of subsection 17.8. in connection with the critique against FTTs' producing liability even under loss-bearing transactions; as a relevant example of such asymmetric treatment, the three authors point out the moral hazard problem, which resulted in the winners' of the recent crisis passing the bill on to third parties, as mentioned towards the end of the introductory paragraph of section 8. Taxing high returns at higher rate is exactly one way of achieving this asymmetry, as argued by Keen et al. (2010). In their opinion, the definition of excess return can be most easily carried out using an accrual-concept tax on capital return, such as a corporate income tax, or allowance for corporate equity – it will simply be the surplus above some threshold rate of return set well beyond "normal". Unlike FAT1 and FAT2, Keen et al. (2010) argue that there ought not to be any loss offsets in the case of FAT3, owing to its purpose to affect incentives. They then argue that at least the systemic institutions should be subject to FAT3, as they are the source of the largest amount of the externality stemming from distorted incentives. The tax on excess return may in principle be levied on all stakeholders of a given entity, including all employees, but unlike FAT2, FAT3 may command different rates reflecting how much the actions of a particular co-worker contribute to the company's risk-taking. NOU (2011:1), however, is sceptical about the possibility of precisely measuring the latter.

From the exposition in Keen et al. (2010), it appears that FAT1 will have the broadest base, while FAT3, the smallest; for instance, IMF (2010:69) provides very rough estimates for a number of countries, based on figures from 2006, where the three bases for Norway as percentages of GDP are estimated to be 2.7, 1.5 and 0.3, respectively. The exact definition of the tax base would determine the tax's incidence and revenue potential; two instances are provided by IMF (2010): the more precisely the tax falls on rents, the greater the reduction of incentives to pass it on to customers, while, on the other hand, levying a FAT on all wages makes it more likely that the tax will be passed on. As far as tax rates are concerned, FAT3's should have the highest, whereas FAT1's the lowest one(s). The rates on FAT2 and FAT3 will be more sensitive to cross-border arrangements, as they are envisaged as being levied on origin basis, with the rate of FAT2's having to be the lower, the less international coordination there is.

A potential problem with all of the proposals is the lack of absolute criteria for the precise determination of what constitutes normal returns. Further, IMF (2010) cautions it may be difficult to distinguish the rent element from returns to high productivity when one assesses excessive remuneration in the financial sector. In addition, the results in De Nicolò (2010) open for FATs' actually encouraging banks to increase their risk, as a worst case scenario. Similar consequences are implied by the results of Huizinga, Voget and Wagner (2011).

19.2. FATs as a complement to, and an improvement over existing arrangements

One area where FATs constitute an improvement of existing arrangements is the mitigation of moral hazard stemming from deposit insurance; more specifically the practice can give incentives to banks to take higher risk than what is optimal from a social point of view, because it provides them with a source of inexpensive financing¹⁰⁸. As argued in the literature, since the difference in costs between the use of deposits and alternative forms of financing is increasing with the risk taken by the bank, deposit insurance is worth more to a financial intermediary which takes high risk than to one which takes low risk and, all else equal, its value increases with the amount of risk taken. A less obvious, but within the scope of the current composition natural way to look at this is to realize that while the insurance is provided to depositors, it also exerts a positive externality on banks. So, a Pigouvian tax or a fee can be employed to make the latter pay for the benefits they receive. More specifically, it will incentivize banks to realign their lending practices along socially preferable lines, by eliminating the discount at which they acquire the “lottery ticket” of providing funds to rather risky borrowers (who can be thought of as the *speculative borrowers* of Minsky). The tax may, in principle, be used to discourage banks from lending to people who have no chance of paying back (who can be thought of as the *Ponzi borrowers* of Minsky), only to cash in for some time and then make use of the deposit insurance (and other forms of government support). However, since the latter issue is the result of the changes in the way the financial sector, and banks in particular, operate, more comprehensive measures are necessary to address it. It is clear that if pecuniary measures are to be effective, the way participation fees in deposit insurance schemes are determined must reflect the actual risk at the particular bank.

Since deposit insurance schemes can be compared to ordinary insurance schemes, and deposit insurance fees to insurance premia, it is in theory possible that banks, if given the opportunity to choose whether to participate in such an arrangement do so in crisis times and opt out in normal times. The committee behind NOU (2011:1) is of the opinion that they should be mandated to participate in a deposit insurance scheme and proposes the use of fees to prevent them from selectively doing so, at least as far as the ones operating in Norway are concerned. The latter issue is relevant primarily for branches of foreign banks doing business in the realm. The reason is that while mandatory participation of the latter in the Banks’ Safety Fund (*Bankenes sikringsfond*) would be optimal for depositors, such a practice runs against EU law. To prevent the branches of foreign banks from speculating against the deposit guarantee scheme by depriving them of the opportunity for opting in and out of the safety fund at opportune times, they should pay non-membership fee. It would fall away once a branch joins the insurance scheme. Entities willing to leave the arrangement should face an opt-out fee, possibly combined with some mandatory termination term, say, two years, as the report proposes.

As mentioned in subsection 18.4., FATs can be levied on dividend payments as a means of addressing the problems caused by the corporate way of organizing an enterprise, where everybody has a share of the spoils when things turn up well, and no one is liable when they turn bad. Thus, they can be used to counter the excessive short-termism and risk-taking of corporate entities. More specifically, FAT3, according to Keen et al. (2010), can complement regulation to meet the latter goal.

Another area where FATs can find application, mentioned in subsection 18.6., is as compensation for the financial sector’s VAT exemption – such taxes, if properly designed, can alleviate the misallocations, caused by its undertaxation – reduction of distortions in consumption and business structure (primarily offset the tendency of the industry’s becoming too large due to favourable tax

¹⁰⁸ The reason is that depositors will accept lower interest rates for any given risk, as long as their deposits are insured, because the arrangement provides them with a de facto risk-free return on the guaranteed amount. An analogy can be made to a gamble, where one retains one’s bet, even if one does not win anything.

treatment) — and, in addition, provide significant income. According to NOU (2011:1), the repeal of the financial sector's VAT exemption through the use of FATs will be more precisely targeted, while in *The Mirrlees Review* (2011) it is argued that such an approach will not affect transactions outside the financial sector on which VAT is levied and/or reclaimed, and will also ensure that its administering is easy. The proposed tax base is the sum of profits and remuneration in financial institutions, in essence resembling FAT1 of Keen et al. (2010), since it can be used as measures of value added by the financial sector. The tax itself may be introduced separately on financial institutions, while the existing VAT system is retained. As *The Mirrlees Review* (2011) argues, such a separate tax raises two issues: the scope for avoidance and, more importantly, the need to ensure its appropriate interaction with the existing VAT framework. As far as the former is concerned, in order that effect of the high mobility of financial services as a major hindrance to the implementability of such a tax be minimized, it (the tax) should be introduced simultaneously in as many countries as possible and with as similar rates as possible. Regarding the second issue one would have to take into account that if the rate(s) of the new tax differ(s) from the one(s) of the existing VAT, clear definitions will be necessary to distinguish real from financial activities, lest vulnerability to lobbyism blunt the effect of the tax — as such a scenario is considered a very real possibility, as the report points out. This problem would not arise if the tax rates are the same, but as noted before, to address the externalities caused by the financial sector, different rates may be necessary. The repeal of the financial sector's VAT exemption through FATs faces some additional challenges, as well. One, according to *The Mirrlees Review* (2011), is that for the implied equivalence of the FAT and the VAT on financial services to hold, the former would have to meet two rather unusual criteria when profits are calculated: first, some workable form of allowing business-to-business customers to recoup their losses would have to be arranged. Second, if total profits and wages form the tax base, then some of the activities would be subject to double taxation — once to an already existing VAT, for instance on provision of deposit safes — an apparent violation of the neutrality principle. Therefore only those profits and wages which arise as a result of VAT-exempt activities should be subject to FAT, reflecting the more general insight that for efficiency reasons it will be crucial to ensure that any cash flow is subject to only one of the taxes. Attaining this latter goal will have the practical benefit of ensuring that the tax base can be easily determined in that any cash flow not subject to the VAT will be subject to the separate tax on financial services. Such “a partition of financial institutions activities” is also advocated by Keen et al. (2010:129) as the most suitable approach to coordinating FAT1 with the existing VAT system.

20. Taxing Variable Payment

The remuneration arrangements in the financial sector, and especially their variable components, have been controversial for various reasons. The recent crisis has added another dimension to this nature of theirs, both on the micro- and the macro-level. On the former it revealed that excessive variable payment, measured both by its wide-spread use and lavishness, can foster increased short-termism and encourage superoptimal risk-taking, if its design is flawed. On the macro-level, the far-reaching negative effects of such practices on third parties demonstrated that remuneration schemes in financial institutions can affect the stability of the financial system, and hence negatively impact the economy, as stressed in NOU (2011:1).

Shaviro et al. (2011) list three reasons for why (what is ex post recognized as) ill-designed remuneration schemes, which are by no means unique to the financial sector, can adversely affect social welfare in such a dramatic way. First, financial firms have the ability and know-how to shift residual risk to third parties; more specifically, derivatives and ingenuity in evading regulatory oversight allow them to bet with limited capital using highly complex and intransparent techniques. Second, even if/when negative tail risks materialize, ex ante explicit commitment to and ex post ad hoc implemented support measures ensure that financial firms and their managers might not suffer

as much as other sectors of the economy do, thus removing incentives to exert caution when risk is taken. Third, on the other hand, with proper incentives in place, financial institutions would actually be more cautious in accepting tail risks than other types of firms because of the danger that even an unfounded suspicion of solvency problems may turn into a self-fulfilling prophecy, which creates sensitivity towards risk both on the micro- (an individual institution's) and the macro- (the financial sector's) level not normally characteristic of other sectors.

The recent experience and the above theoretical explanation for it suggest that, as part of the overhaul of regulation aiming for a more robust financial sector, the risk posed by remuneration arrangements in financial institutions should be accorded due attention, because, as it is argued in NOU (2011:1), there is no reason to treat it differently from any other risk. More specifically, the implemented changes should counter the effects of wage schemes on incentives to take excessive risk and to be overly focused on the short-term, both by individual employees and firms in general. Thus, according to the report, the design of the new framework should incorporate the general lesson from the recent crisis that risk-taking will always be able to find its way to markets and institutions with minimal regulation, and the ingenuity of financial firms when it comes to adapting to regulatory changes so that their profits do not suffer. This implies that the formal meeting of requirements will not be enough. Instead, the supervisory authorities should follow closely the development and regularly evaluate the actual incentive effects which can follow from the way financial institutions remunerate their employees.

In reaction to the crisis some countries have introduced levies on variable payment. These measures mostly cannot, in my opinion, be considered as some serious attempt to address the misallocations which incentive-improper remuneration arrangements are blamed for. Rather, they can be regarded as nothing more than ad hoc eyewash meant to appease a public aroused at the outrageous benevolence shown to the financial sector by policy makers. It remains to be seen, whether there will be general genuine interest in dealing with the issue at all. As of today, there seems to be lack of political will to address the problem. And this is not just my impression: von Ehrlich and Radulescu (2012:2) provide the following quote from L. Zingales' "The Bonus Risk" (2010, Project Syndicate): "If the solution is so simple, why has no elected body implemented it? My fear is that politicians want to be perceived as tough on bankers, but have no interest in really fixing the problem".

20.1. The adverse impact of bonuses on incentives

Primarily NOU (2011:1), von Ehrlich and Radulescu (2012), and Shaviro et al. (2011) have searched for the causes of bonuses' having undesired effects (from society's point of view) on the incentives of the actors in the financial sector. The exposition in this subsection is a synthesis of their findings.

First, for bonuses to have such a bearing on risk-taking and short-termism, they must have comprised a significant part of the remuneration in the financial sector. This, indeed, has been the case. In the literature the high proportion of variable to fixed pay has been more or less identified as the prerequisite feature of wage schemes' impact on incentives to engage in such unfortunate practices.

Second, as part of the general trend of shrinking horizons, the focus was to deliver results in the short run. This was reflected in the remuneration arrangements' making heavy use of short-term cash bonuses which rewarded performance measured by volume and marked-market profits, rather than the long-term profitability of the investments. One example is provided by Shaviro et al. (2011) who argue that wage schemes designed to exact maximum performance from managers in order to generate extraordinary returns provided them with economic incentives to follow what the authors call "nickels in front of a steamroller" strategies — they return profits in excess of what can be viewed as normal most of the time, but occasionally lead to huge losses —, even if the expected return to shareholders was negative.

Third, the short-term orientation of the variable pay schemes contributed to a considerable risk asymmetry, as the maturity of the investments often was long, compared to the attachment to the firm of the ones managing them.

Fourth, appropriate adjustment of results for risk was the exception, which often entailed that the remuneration schemes had limited or no downside for the particular employees. This was especially problematic in view of the fact that few managers could fear being held responsible for ex post bad decisions, as long as they could convince others that ex ante their choices were well-founded in business judgement.

20.2. The necessity for regulating bonuses

Within the scope of the current composition, the main rationale for regulation of variable pay in the financial sector is the correction of market failures it is believed to cause, or at least is strongly conducive to. With the “benefit” of the crisis we can point out some of the forms they take. Von Ehrlich and Radulescu (2012) identify four¹⁰⁹ and mention some remedies to them.

First, the presence of asymmetric information between bondholders and managers to the disadvantage of the former allows the latter to engage in excessive risk-taking. The reason is that managers are hired to increase shareholder value, that is, they are agents whose principal are equity holders, and it is an established result in the literature on corporate finance that the latter are more open to risk than debt holders are. The remedy proposed for this issue is to base compensation on the price of debt, for instance in the form of credit default swap spreads.

Second, the implicit and explicit protection against downside risks enjoyed by investors, for instance because of limited liability as Shaviro et al. (2011) point out, is conducive to their being willing to tolerate more risky behaviour by the ones in charge of their money. To deal with this issue a combination of regulation on bonus structures and a tax on their level appears appropriate.

Third, the competition for bankers produces a negative externality, because it drives up bankers’ wages and thereby implicitly the default risk of rival banks. This issue cannot be resolved through taxation of bonuses, because even if the fraction of variable pay were to be lowered, default risk for banks will remain unaffected.

Fourth, the bonus component can be employed as a screening device to separate high- from low-ability managers. However, its price takes one form in the induced superoptimal risk-taking, both from a micro- (an individual institution’s) and macro- (society’s) perspective.

20.3. Practical difficulties

According to von Ehrlich and Radulescu (2012), the empirically confirmed positive correlation between performance-based payment on the one hand, and volatility of returns and idiosyncratic risk profiles of firms on the other support the idea that reducing variable compensation should disinvite short-termism and preference for risky strategies, but only when this part of the salaries is considered *in general*. However, once a distinction is made between equity compensation and cash bonuses, it is so that the former is the driving force behind risky behaviour, while the latter have no significant effect. The authors suspect stock options to be the main culprit, because they put a limit on downside risk to managers, but not on upside gains, thus inducing risk-taking, even without a prospective bailout. Based on this finding, von Ehrlich and Radulescu (2012) conclude that the UK bonus tax on the financial sector in 2009/2010 was poorly designed, and might have even had

¹⁰⁹ The two authors note that one thing all these four issues have in common is that different types of performance-based pay are not distinguished. The variable compensation component is normally linear, hence the link between various specific components and risk cannot be captured. Neither can be the income shift when only some components of variable pay are taxed.

adverse effects on corporate risk-taking¹¹⁰. They explain the latter with the results of their paper, which show that equity compensation actually encourages risky behaviour, so that a shift towards such kind of variable pay — which could be the case if only cash bonuses (which have received virtually all negative attention) are taxed — may have adverse effect on corporate risk-taking from society's point of view. They argue that to reduce the incentives to engage in such actions, the tax base must be extended over variable equity compensation too, because simply changing the payment structures of executives does not improve risk profiles and stability of firms.

Keen (2011) argues that taxes on bonuses will inevitably involve some selectivity. The majority of the committee behind NOU (2011:1) is of the opinion that this is not case for systemic institutions. Instead, the rules for remuneration applying to the latter should be expanded to include all employees, because the delimitation of which ones among them have a say on the institution's risk exposure may open the door to unwanted adjustments, all the way to effectively nullifying the intended effects of regulation.

Shaviro et al. (2011) argue that there are no features of a bonus tax which are likely affect in a socially beneficial way the incentives facing the actors in the financial sector and note that such a tax will easily be avoided and may be rolled over to shareholders.

20.4. Empirical findings

Von Ehrlich and Radulescu (2012) studied the effect of the bonus tax imposed on the UK financial sector for the period 2009/2010 and risk-taking, finding no evidence of a significant impact of bonuses on the latter. They also stress that the tax was poorly designed, which allowed its avoidance: cash bonuses were reduced, while other components of pay were increased, so that the executives' overall compensation was unchanged – thus, it seems to me that the tax fell prey to a weakness of FTTs, discussed in subsection 17.8.: the fact that there are multiple ways of arriving at a particular economic outcome, which need not entail the same tax liability. The resulting higher proportion of equity- and/or stock-based variable compensation might have had an adverse effect on risk-taking. The lesson to be drawn from the UK experience is that if the taxation of bonuses should achieve the intended goals, proper setup is a prerequisite.

20.5. Alternatives

Instead of taxing variable payment, NOU (2011:1) proposes the imposition of additional capital requirements on financial institutions, the remuneration arrangements of which, in the opinion of the supervisory authorities, invite too much risk-taking.

21. Bank Taxes

The third group of proposed corrective taxes and fees — that of levies on financial institutions' liabilities — is exemplified by bank taxes and stability fees. The two share common intuition and similar aspects. In general, taxes and fees based on the balance sheets of financial institutions can be suited to obtain some of the same goals as capital requirements like, for instance, limiting risk-taking and borrowing with respect to equity. As IMF (2010) and NOU (2011:1) inform, this form of fees is already in use in connection with deposit insurance, but its area of application may well be extended in the future. In countries where it is levied (for example, UK, Sweden, Germany), the declared goal is

¹¹⁰ Weder di Mauro (2010b) also mentions this tax as one of those that served other purposes than ensuring a more robust financial sector, because they were not designed to reduce systemic risk of the global financial institutions. More generally, Keen et al. (2010:134) note that the implemented taxes on bonus payments focused too narrowly on certain forms of wages only, and the benchmarks used to determine excessive remuneration had been set in a rather arbitrary manner.

to charge financial institutions for their contribution to systemic risk and to collect funds on which to draw means to support troubled entities in future crises (Germany and Sweden) or revenue-raising (UK)¹¹¹. The focus in this section will be on bank taxes.

As NOU (2011:1) argues, the recent crisis demonstrated the importance of bank capital's ability to bear losses for the stability of the financial system (and exposed the inadequacy of the then-in-place capital arrangements' ensuring the sustainability of liquidity positions of many financial institutions, as well, with market finance's, especially short-term, having been identified as the prime danger to financial firms' liquidity). As a result, measures — new capital and liquidity requirements — have been taken in consideration to ensure that sufficient buffers are built up during good times, so that banks are able to withstand episodes of distress of severity similar to the latest one, or at least are able to continue normal operation in relatively short spells of trouble, without being dependent on government support, correcting the pre-crisis arrangements' not making it privately optimal for a given institution to store enough capital for a rainy day¹¹². This is especially important for systemic institutions, the existence of which makes the case for a differentiated approach regardless of the measures taken.

While the focus of the new arrangements, as has been the case with other issues, has been on improved regulation in the form of stricter capital adequacy requirements, owing to the approach's history of use and ease of implementation in practice, as pointed out in Korinek (2011a), there is the alternative approach through bank taxes, the employment of which is, according to Keen (2012, slide 11), "an important experiment". The scope for their effective use is provided by asymmetric information: private knowledge by the banks implies a difference between privately optimal and socially optimal capital ratios. They also offer some advantages over capital requirements. Firstly, taxes are "a more dynamically efficient instrument" (IMF 2010:51), which means that they have a smoother impact on bank finances, by imposing continuous costs, so that the risk of, for instance, deleveraging or costly raising of new capital is reduced. Secondly, the report argues, they offer more flexibility during a crisis.

The envisaged form of implementation is as a tax on uninsured liabilities, which, according to Shaviro et al. (2011), is a substitute for capital requirements for the regulatory purpose of reducing bank risk, and the two should therefore be coordinated. The idea behind taxing uninsured liabilities is to reduce the incentives banks have to use unstable funding to finance risky lending and investments. In addition, bank taxes will promote the build-up of buffers and counter the debt bias. As far as systemic institutions are concerned, NOU (2011:1) argues that special taxation can be used as a means of "compensating" this kind of entities for the lack of risk for their management and owners, as they operate in practice with an implicit guarantee. In order that this measure be effective, the taxes must have an effect on the leverage and capital ratios of banks. Keen (2012) and IMF (2010) argue that the practice of holding buffers suggests that there is scope for some effect, but how strong is it? According to the former source, theory postulates that the risk of violating capital requirements makes leverage less responsive, while flatter agency costs, which account for high leverage in the first place, imply greater responsiveness. This undefinedness means that the answer must be searched for in empirical studies. The results of one such encompassing 14,000 banks in 82 countries over the period 2001-09, mentioned by Keen (2012, slide 8), are that taxes do affect leverage of financial intermediaries, by as much as is the case for non-financial firms and more quickly, but the effect on large banks is much smaller. Further, Keen and de Mooij (2012) find that banks with large capital buffers are very responsive to taxes, but not to capital requirements, while the inverse is the case for the ones sitting on limited reserves. Importantly, they also find that the

¹¹¹ The arrangement is not perfect, however, as NOU (2011:1) points out. For instance, currently within the EU, differences among such taxes in rates, purposes and design can lead to distortions and suboptimal adjustments in the internal market. Further, some banks may end up being subject to double taxation.

¹¹² For banks the more suitable term in view of the experience during the crisis would be "a deluginous day".

largest five percent of the banks, holding almost sixty percent of all bank assets, are considerably less responsive to taxation and are unresponsive to minimum capital requirements. This reduced responsiveness to taxation and virtual insensitivity to capital requirements may be used as a justification for the use of non-linear taxes. Another one is provided by Shaviro et al. (2011): the bank tax should be modelled like a capital requirement to increase with the riskiness of the financial intermediary. The reason they give why its rate would have to be risk-weighted rather than flat is that as an optimal corrective tax it would have to reflect the marginal social damage which a given institution's actions could cause. The dynamic nature of financial markets necessitates that the factors used in the determination of the tax rate be measured virtually continuously, so that it is up-to-date as often as possible. A serious challenge, however, is the difficulty of accurate and transparent measurement of the risk factors in practice.

21.1. Concrete proposals

One concrete proposal, mentioned in NOU (2011:1), has been put forward by the British government. According to it, the balance sheets of all banks doing business in the UK are to be taxed. Besides raising revenue, the tax is meant to reduce the incentives of financial intermediaries to rely on short-term market financing by subjecting financing of over a year's maturity to a lower rate.

Another proposal, already mentioned in subsection 9.4., due to Shin (2010) is to tax non-core liabilities of banks. His definition of core and non-core liabilities has already been presented in footnote 48. The author clearly identifies its role as an incentive-correcting device. As mentioned in subsection 11.2., Matheson (2011) comes with virtually the same proposal – see the next-to-last paragraph in subsection 11.2. and footnote 77 for his idea and its difference from Shin (2010) respectively. Keen and de Mooij (2012) also mention taxes on banks' wholesale borrowing as a remedy to the challenges to policy design they identify. According to Shin (2010), a tax on non-core liabilities has several advantages: it is easy to implement, less distortive, can be targeted at the activities causing the greatest spillover effects, and is not likely to be avoided easily. Further, its suitability for both advanced and emerging economies should facilitate its imposition on the global level. Last, but not least, it is a price-based intervention, which means that it works with the market forces and (God forbid) not against them, to the relief of free-marketeers.

Despite the aforementioned advantages, Shin (2010) argues that to maximize the effectiveness of the tax and to minimize the scope for distortions through possible circumvention or shift in capital flow patterns (but not the scale thereof), its introduction should be globally coordinated and it should charge the same rate. That notwithstanding, a country or a group of countries may, in principle, nevertheless decide to introduce a tax on non-core liabilities unilaterally and still reap substantial benefits from dampening the financial cycle, on condition that no distortions arise from attempts to avoid the tax. The latter can be achieved through coordinated use with other tools of prudential regulation; for instance, membership in a settlement system can be used as a criterion for who should be subject to the non-core liabilities tax, because someone along the chain will end up paying it — reminiscent of the proposal in NOU (2011:1) to charge non-members of *Bankenes sikringsfond* a fee, mentioned in subsection 19.2.

To strengthen the case for his proposal, Shin (2010) compares how a tax on non-core liabilities fares against other measures which may be employed to dampen the procyclicality of the financial system and the associated problems, and concludes that the alternatives appear to be inferior as tools of macroprudential regulation.

First, he dismisses time-varying capital requirements as a viable option on the grounds that the recent crisis demonstrated that they do not bind sufficiently during a boom – banks ended up with too large balance sheets and too little capital to meet the downturn. Their improved version — time-varying capital requirements which try to reflect the stage of the financial cycle more precisely — requires a lot of measurement and raise various technical issues. Further, there is the possibility of

bureaucratic delays in implementation. All these issues speak against time-varying capital requirements as something of practical significance

The next measure — the practice of expected loss provision, pioneered by the Bank of Spain — had success in the recent crisis, but it, too, is subject to the same measurement and technical issues when it comes to calibrating for the amount necessary. Closely related is the question of how severe a downturn the expected loss provision will have the capacity to meet. Lastly, there is the time-inconsistency problem: history has shown that when trouble comes, supportive measures of any proportion will be implemented anyway, removing the disciplining effect of the measure.

The last alternative he considers is administrative command, more specifically a cap on the loans-to-deposits ratio of a particular bank, which is equivalent to a tax with a rate of zero up to the mandated threshold, and of infinity above that. The measure serves the purpose of tying the growth of lending to the size of the deposit base¹¹³. The drawbacks of the practice are its inflexibility and the potential for price distortions in financial markets, in the form of increased interest rates on deposits because of competitive pressure. In turn, the gap between deposit and market interest rates will strengthen the incentives to engage in circular financing where market finance is ensured through bond issue and the proceeds are then deposited in the banking system; Shin (2010) cites the experience of Japan in the 1980s which led to the bubble's there being exacerbated as an example.

21.2. Incidence

None among Shin (2010), Shaviro et al. (2011) and Keen and de Mooij (2012) considers the incidence of bank taxes. On the other hand, NOU (2011:1) provides a thorough analysis of the incidence of higher capital requirements. To the extent that they and bank taxes can be considered substitutes in their effect on the costs of financial institutions, one could use results of the report as a rough proxy to determine the incidence of the latter policy instrument. The intuition is correct, indeed – Weder di Mauro (2010a:96) argues that “the determinants of the ultimate incidence are the same for both types of regulatory intervention [capital requirements v. taxes]: faced with higher costs the extent to which banks reduce the return on equity or increase the interest rates on loan [*sic*] will depend on the market structure and on the competitive position of the bank in this market”.

According to NOU (2011:1), the imposition of higher capital requirements on banks will likely increase their costs, as it is normally assumed that they pay higher risk premium for equity than debt on the margin. Further increases in costs will be brought about by more long-term financing and greater liquidity buffers. The committee behind NOU (2011:1) believes that much of these costs will be compensated for by savings in the form of lower interest on banks' borrowing, because subject to the new rules they will become less risky borrowers. Over the medium term it will be the competition in the banking sector which will determine how much of the increased costs will be passed on to customers and how much will be met with lower profitability for banks and their owners. With intensive competition, there will be low profit margins and hence little room for absorption of increased costs, implying that a higher proportion of potential increased outlays is passed on to customers. Further, the division of expenses among the banks will play a role as well: increased costs which affect all banks equally will be passed on to a large degree. Similar conclusions arise as an implication of the results of Huizinga et al. (2011), who focus on the wider taxation of banks.

¹¹³ This is akin to the nowadays not very common practice of central banks' controlling reserves as an instrument of monetary policy.

22. Stability Fee

One of the measures to correct the market failure and promote the stability of the financial sector which has been discussed is the imposition of what is called a stability fee, or a “Financial Stability Contribution”, as it is referred to in IMF (2010) and Cottarelli (2010). Its proponents argue that, properly designed and implemented, it will align the incentives of the actors in the financial sector, so that they internalize systemic risk, giving them incentives to reduce it. In addition, of marginal interest within the scope of the current composition, it is a potential source of funds for state intervention in the financial sector. Nevertheless, to the extent that the proceeds may be used to pre-fund future support measures, which within the new framework of improved resolution mechanisms are expected to lead to greater financial stability positively affecting all financial institutions, the fee may be regarded as a means of charging them for the benefit they draw from what is in effect a positive externality, thus helping to align private with social incentives.

The stability fee is regarded as a substitute for the prevalent practice of using capital requirements, and, by transitivity, to bank taxes, as well. According to Weder di Mauro (2010b), the dominance of the approach can be explained with experience with application and capital’s remaining on the balance sheet of banks. Capital requirements have their weaknesses, though, two of which have been pointed out in Weder di Mauro (2010b) and Weder di Mauro and Klüh (2010). First, even under the new framework of Basel III, they will be used for too many goals, not necessarily compatible with one another¹¹⁴. Trying to achieve too many goals with a single instrument “will inevitably involve trade-offs, lead to a system of capital requirements, which is highly complex, intransparent and prone to manipulation, constant re-interpretation and forbearance” (Weder di Mauro 2010b:3), thus violating all four previously mentioned criteria listed in Weder di Mauro (2010b) as essential for any policy tool. Second, a persistent problem which has shown up during previous crises is that capital requirements apply only to de jure banks, while systemic relevance is not limited solely to them, so that, unless complemented with other measures which affect the non-bank systemic institutions, they only provide a partial solution, and may actually encourage further migration to the less regulated venues of the shadow banking system.

22.1. *Design*

What will comprise the “base” of the stability fee? It should be based on balance sheet measures, because, compared to other variables, they provide the clearest picture of a financial institution’s risk. More specifically, according to NOU (2011:1), Weder di Mauro (2010b) and IMF (2010), it should consist of all liabilities net of equity capital (to prevent adverse effects on capital accumulation) and insured deposits (and other liabilities for which reasons for exclusion¹¹⁵ can be found¹¹⁶); the third of the aforementioned sources argues that the base may consist of specific liabilities, if one aims at reducing the incentives to engage in the corresponding activities, but this would heighten the risk of unintended effects. The same reference also opens up for the inclusion of some off-balance sheet items, to the extent that they may be a source of systemic risk, but argues that issues regarding off-balance sheet instruments would be primarily addressed through improved direct regulation. IMF (2010:53) rationalizes the basing of the fee on liabilities rather than on assets on the fact of liability risks’ being “easier to capture consistently across banks [or other financial institutions]” when supervisors operate with insufficient information. In addition, the practice would provide some correction to the preferential treatment of debt for corporate tax purposes. It also provides for lower rates of the fee, reducing the scope for “unintended distortion” (IMF 2010:13) and, lastly, reflects the

¹¹⁴ One current example they provide is the practice of using them both as a buffer against unexpected losses and as way to reduce risk. Proposed additional uses include control of liquidity risk and mitigation of procyclicality.

¹¹⁵ IMF (2010) argues that allowance for insured liabilities is a better option than exclusion thereof.

¹¹⁶ See footnote 54 in IMF (2010:56) for candidates and rationales.

need to support liabilities in the event of a crisis. IMF (2010) argues further that the base should vary with the type of institution, reflecting different contributions to risk; for example, the base of insurance companies, owing to the lower volatility of their funding, should be smaller than the one for banks.

All the authors who study the topic agree that the effectiveness of the fee requires that it be differentiated, which stems from the need to reflect the contribution to systemic risk by each individual firm. Therefore, the fee amount/rate (depending on how one decides to implement it) should vary with the size of the externality each institution is deemed to cause. The variation can be quite complex: for instance, Acharya et al. (2010) argue that the costs of the contingent capital insurance (which determines the size of the fee in their proposal) should increase non-linearly, “the stronger the capital requirement and the higher the [positive] correlation between the firm’s equity return and the market’s return” (Acharya et al. 2010:11¹¹⁷). In addition, they note, these factors also interact non-linearly. Nevertheless, IMF (2010) argues that in the beginning, the rate of the fee will have to be flat, yet varying with the different types of financial institutions, and only later is it to be refined along the lines suggested by other authors, implying variation within each type. In addition, according to the same reference, it “should reflect differences in the structures of financial systems” (IMF 2010:12). NOU (2011:1) notes that it is desirable that the fee also reflects the potential expectations of state intervention, while the Weder di Mauro (2010b) argues that the fee amount should be set such that it eliminates the implicit subsidy enjoyed by (systemic) institutions altogether – thus, the stability fee offers some remedy to the systeminess bias I have written about earlier; IMF (2010:54) is more explicit on this point: “[l]evies can be set so as...to reduce...incentives to become TBTF [too big to fail]”. More specifically, Weder di Mauro (2010b) proposes to base the fee amount/rate on a risk score, which is to be produced by a formula into which data on indicators such as size, interconnectedness, complexity, maturity of liabilities, etc., is fed¹¹⁸; similarly, reliance on both quantitative and qualitative indicators has also been made the case for in IMF (2010). The risk scores are to be updated at regular intervals: on a yearly or quarterly basis, for instance. As Weder di Mauro (2010b) argues, they confer two additional benefits: they will provide feedback whether the tax works and can be used as an early-warning device. It remains to hope that these risk scores will have more success in reflecting the conditions they attempt to quantify than (have) do(ne) credit ratings.

Acharya et al. (2010) propose the imposition of a stability fee on each institution in the form of a systemic risk tax, which will be determined by the cost of the contingent capital insurance the entity will have to purchase. The tax on each firm will consist of two components: the expected loss of defaulting on liabilities guaranteed by the government, and the expected systemic cost times the percentage contribution by each firm to aggregate financial sector losses above a certain threshold (which the authors argue can be regarded as the “contribution” by each institution to the possible collapse of the financial sector), with the former term determining the level of the tax, and the latter determining which institution pays more of the levy.

The proposal by Kocherlakota (2010) to use “rescue bonds” mentioned earlier is similar to the previous one. The fee charged by the government should equal the reciprocal of the fixed fraction of the expected discounted value of the transfers to the stakeholders of each firm times the price of the bond.

¹¹⁷ The version of their paper I have used is not paginated, so the quote is from what is the eleventh page, including the title page with the abstract of the article in the count.

¹¹⁸ The proposal to use a risk score is based on the author’s scepticism about the ability of the various theoretical models put forward to measure systemic risk and map the possible channels of propagation to produce precise results in practice; a similar view is expressed in IMF (2010).

Lastly, Jeanne and Korinek (2010) argue that within the model which they develop, requiring banks to pay into a crisis resolution fund is in effect equivalent to an ex-ante prudential tax, provided that the amount due from each financial intermediary is set proportional to its contribution to systemic risk.

IMF (2010) argues that the fee should have a broad perimeter, ideally levied on all financial institutions, for one charging them for the positive externality in the form of greater financial stability, but also because if imposed only on a narrow group of institutions it would encourage the migration of systemic risk, and may create scope for moral hazard, whereby the targeted institutions would be implicitly singled out as less likely to fail.

Big and complex institutions pose additional challenges because of their pivotal role in the financial amplification process. The stability fee should, therefore, have stronger impact on their behaviour. This may necessitate coordinating it with other measures specifically targeted at systemic institutions. Indeed, their importance would also necessitate the creation and constant updating of a list of such institutions.

22.2. Issues to take into account when designing the fee

Weder di Mauro (2010b) and Shaviro et al. (2011) stress the importance of preventing the flight to untaxed venues (not just in the geographical sense), if the fee is to achieve its goals. One necessary step to ensure this is the incorporation into the regulatory framework and supervision of the recognized fact that there is a wide variety of what can be termed “financial activities”, the specific realization of which will be to levy the fee on activities, not on types of institutions, so that any institution which may need support should be liable, regardless of its de jure type. This implies that the financial subunits of non-financial firms should be brought under the fee. It should also be derivative-proof, so that its intended effects are not but blunted by the replication of positions through exempted derivative products (as may happen with FTTs, and has happened with poorly structured taxes on remuneration).

There does not seem to be any agreement on the cyclicity of the fee. On the one hand, Shaviro et al. (2011) argue that it will depend on the state of nature, lest it impose further liabilities on the micro- or the macro-level exactly when one or several firms, or the financial system are fragile. On the other hand, Weder di Mauro (2010b) argues that the systemic risk levy should not be varied over the cycle, because its target is tail events¹¹⁹. IMF (2010) argues that one of the refinements of the fee would have to reflect variation of overall risk over time, so that the rate of the fee would have to be adjusted, with the aim of reducing the financial system’s procyclicality.

Primarily Shaviro et al. (2011) argue that international coordination is desirable, because externalities can easily cross borders, as the Asian crisis, the recent financial and sovereign debt crises have shown. While the effect of such a tax will be greater, if it is introduced in the same or similar forms in as many countries as possible, local imposition can be considered, provided the implementation is robust against avoidance.

Weder di Mauro and Klüh (2010) caution that even if the fee manages to reduce the amount of systemic risk “produced” by financial institutions, especially systemic ones, for some of them the private optimum will still involve remaining a threat to financial stability in case of distress. Dealing with the issue requires that proper resolution mechanisms for failing/failed (systemic) institutions be set up.

¹¹⁹ She believes that the procyclicality of the financial system can be reduced by the use of other instruments, such as contingent capital.

The results of De Nicolò (2010) suggest that, as is the case with FATs, the reduced asset growth, net profitability and more expensive capital formation, which the stability fee will entail, may in the worst case lead to banks' increasing their risk.

Last, but not least, it is pointed out in the literature that the necessity to correct the incentives of financial institutions is a completely separate issue from the one whether they are paying the expected value of the insurance coverage on their liabilities (as the latter aims neither at reducing the likelihood, nor the wider economic magnitude of future crises). Thus, all the authors who consider the creation of a resolution fund filled with the proceeds of the fee stress that the latter should continue to be levied, even once the fund has reached its mandated size.

22.3. Incidence

The question of incidence is important in view of the need to limit the scope for nullifying the effect of the fee, minimization of efficiency losses and other unwanted distortive effects. Considerations of cost development and competitiveness will be central in analyzing the consequences of introducing such a corrective measure.

Weder di Mauro (2010b) argues that it is essential that financial institutions not be able to pass the fee on to their creditors and customers; neither should it lead to increased cost of credit to the real sector. She believes that competition between systemic and non-systemic institutions should be enough to achieve this, because it will erode some of the profits, while the levy will make excessive leverage unattractive. NOU (2011:1), however, is sceptical about this – the committee behind the report believes that the financial sector will be able to pass any fees on to its customers, for reasons presented in subsection 21.2., with a minority of the authors of the report even citing this as a main argument for the lack of need of introducing such a fee in Norway for the time being.

22.4. The proceeds

The authors of the various papers reviewed in this section agree that the (primary) purpose of the stability fee is not to raise revenue. Therefore, it is a separate question what will be done with the proceeds from it. They can, of course, go to general taxation and either be used to fund something useful, or, as suggested in *The Mirrlees Review* (2011), to reduce the burden of taxation elsewhere. Primarily Weder di Mauro (2010b), but also NOU (2011:1) and Shaviro et al. (2011) argue that the proceeds can be used to fill a specially created (crisis) resolution fund, the purpose of which will be to reduce the residual (fiscal) risk for the public sector, thus dealing partially with the equity issue: the taxpayers will be spared at least for the direct fiscal expenses of supporting institutions in distress; in addition Weder di Mauro and Klüh (2010:91) believe that “[a] combination of fee and resolution fund would constitute an integrated framework which would allow the public to effectively discipline systemic institutions”. Acharya et al. (2010) propose that the proceeds be put into a systemic risk fund instead, with a view to covering the systemic risk costs. The accumulated means are to be used to support the affected real sector and solvent financial institutions, instead of bailing out failed firms – the authors stress that government guarantees in the form of bailouts to come must be priced separately, a point which has also been made in IMF (2010).

A special-purpose fund is not without conceptual problems, however. A few of them have been shed light on in NOU (2011:1). First, putting money into a special-purpose fund can lead to suboptimal governance of the overall resources of society. Second, the fund, especially if it is big, can reinforce the expectations that the state will intervene, so that it invites for higher risk taking, thus nullifying the effect of the fee. On the other hand, the existence of a fund earmarked to be tapped in the event of a crisis can create the impression that there is some upper limit on how big an amount is available for potential support measures; the existence of a “contingent credit line” (IMF 2010:14) can help alleviate this problem. Further, building on their bailout irrelevance (or, bailout neutrality, as it is referred to in Korinek (2011a)) result, Jeanne and Korinek (2010) argue that a crisis resolution fund

has no effect on equilibrium outcome and welfare in their model, because of Ricardian equivalence: the private agents will borrow more and exactly the amount that will leave them with the same endowment in the bust in the social-planner setting, as in the laissez-faire case.

Lastly, as far as the amount raised is concerned, Shaviro et al. (2011) believe that it may well exceed the potential bailout payments, whereas Weder di Mauro (2010b) does not think that this is likely to be the case in practice. In light of the recent crisis¹²⁰, IMF (2010) points out that while the resources accumulated in a designated fund may suffice to cover actual support measures, they almost surely will fall short of the amount necessary to be pledged quickly around the time trouble begins; the difference will have to be provided by a “contingent credit line” on the part of the government, for which the latter would charge a separate fee.

23. Other Issues

Some final issues, mostly building on NOU (2011:1) due to its comprehensiveness, are in order.

23.1. International coordination

The fact that externalities can easily cross borders makes international cooperation, if not imperative, at least highly desirable. In the likely case that a corrective tax or fee, of whatever type it may be, is not globally accepted, it should be robust to imperfect international coordination. Still, joining the ranks of cooperating jurisdictions can be facilitated by the fact “that no one should think themselves immune from failures and crisis [*sic*]” (Cottarelli 2010:77). According to IMF (2010), international coordination of efforts would be particularly beneficial in dealing with cross-border financial institutions. Cooperation “does not mean identical application everywhere” (IMF 2010:22): for instance, the report suggests that countries may agree on a common tax base and some minimum tax rates, which they are then able to increase, to reflect particular considerations and experiences; indeed, this has already been the case for Norway in the EEA context, as pointed out by NOU (2011:1), and there is the EU practice of setting minimum standards and requirements, leaving each member state the leeway to make them stricter. On the other hand, limited or no participation in internationally-coordinated measures carries the risk of paving the way for future crises.

23.2. The adaptability of the financial sector

A major lesson from the recent crisis is that actors in the financial system are quick to adapt to a changed operational framework and market conditions. The results of such an adjustment are not easy to foresee. This would require continually evolving regulatory arrangements. This point has been made in the non-mainstream literature, and stressed particularly in Minsky (1982; 1986), whereas in the post-crisis literature it is featured in NOU (2011:1).

23.3. Do not put too many eggs in a single basket

The recent crisis has revealed the necessity of regulatory changes which aim at limiting the opportunities for big engagements, that is, some restrictions should be imposed on financial institutions on how much they can loan to a particular borrower or a group of similar borrowers. The idea is to avoid potentially big losses to a financial firm when some of its major borrowers are in trouble – consider the cases of lenders to subprime borrowers and Cypriot banks’ exposure to Greek government bonds as recent examples.

¹²⁰ See Appendix 1 in IMF (2010) for the account.

23.4. Tax havens

Due to its comprehensiveness, NOU (2011:1) looks at the role of tax havens. Their existence and mode of operation raise many issues. The recent crisis highlighted one of them: they can contribute to financial instability, because the large volume of financial capital residing in or flowing through them can adversely affect the scope and efficacy of regulation both on country and international level. More specifically, the report cites an earlier NOU (2009:19) on their “contribution” to the recent crisis:

“Tax havens raise the risk premium in international financial markets. The financial crisis has shown that many financial institutions had obligations which had not been entered in the books, where part of the obligations had been registered in tax havens and thereby shielded against scrutiny¹²¹...Tax havens contribute to increased counterparty risk and greater informational asymmetry between different actors, which weakens the manner of operation of the international financial market and contributes to higher borrowing costs and risk premia for all countries.” (NOU 2011:1, p. 170)¹²²

The committee behind the cited report NOU (2009:19) opines that the work towards thwarting tax evasion should be of high priority both on the national and international level. More specifically, new rules regarding when a legal entity should be considered residing in a closed jurisdiction, new rules for tax law between states and an international convention which aims at preventing that states develop closed structures which can be suitable to cause losses and damage in other jurisdictions must be implemented.

While the fight against tax evasion is important in its own right, the aspect of importance within the scope of the current composition is the prevention of circumvention of corrective taxation, something which according to Schjelderup (2011) tax havens enable. For instance, a corrective tax for holdings and/or registrations in such entities may be levied: just owning an account or being registered there may imply a tax liability.

23.5. Backward-looking taxes

Shaviro et al. (2011) discuss briefly the merits of backward-looking taxes. Examples of such are bail-ins¹²³ — for instance the Cypriot one was sold as a one-time tax, although this is too euphemistic a term for it — and the one-time tax on bonuses implemented in the UK. Such taxes are inherently non-distorting with regard to future decisions, but they do affect expectations about the future course of policy. In addition, opportunities for avoiding such a tax are limited — mostly access to inside information — which makes concerns over international coordination largely irrelevant. The potential loss of credibility (or whatever is left of it) is a major concern, as also IMF (2010) warns, especially in view of the fact that the recent crises (the financial and the sovereign debt one) have shown the enormous pecuniary cost of policy makers’ time-inconsistency. A side matter, mentioned

¹²¹ NOU (2011:1) cites NOU (2009:19), according to which the positions of international banks in tax havens had increased in the years before the recent crisis, and approximately a quarter of the liabilities of banks in countries reporting to The Bank of International Settlements are placed in such jurisdictions.

¹²² A rather literal translation of the Norwegian text:

Skatteparadisene øker risikopremien i internasjonale finansmarkeder. Finanskrisen har vist at mange finansinstitusjoner hadde forpliktelser som ikke var bokført i regnskapene, der en del av forpliktelsene var registrert i skatteparadis og dermed skjernet for innsyn. (...) Skatteparadisene bidrar til økt motpartsrisiko og større infomasjonsasymmetri mellom ulike aktører, noe som svekker virkemåten til det internasjonale finansmarkedet og bidrar til høyere lånekostnader og risikopremier for alle land. (NOU 2011:1, p. 170)

¹²³ Bail-ins have their advantages in achieving the goal of a more robust financial sector, but their controversial nature and questionable resemblance to “proper” taxes make a thorough discussion on them and their role something well beyond the scope of the current composition.

in Shaviro et al. (2011), with reference to the funds mentioned in subsection 22.4., is that an appropriately designed Pigouvian tax will raise a given amount of revenue causing less distortion than a backward-looking tax.

24. Conclusion

In the current composition I have made an attempt to present a synopsis of the potential role corrective taxes and fees can play in the process of overhauling the regulatory framework within which the financial sector operates, in order that the latter be robust and serve the rest of the economy in a better way than it did in the years before the crisis. What conclusions can be made from the reviewed literature?

First, I argued that in the work towards a more stable financial system emphasis should be put on pro-active measures, that is, the focus should be on the prevention of crises and the limitation of their damaging impact, whenever and to the extent the latter is possible; there is agreement in the literature that this can benefit all economic agents. An apparent and necessary pre-condition to achieve these ends is that one has adequate knowledge of the nature of financial disasters – something modern mainstream economics has so far largely failed to deliver. For this reason, I have found it justified to provide a brief survey of certain non-mainstream contributions to explaining the observed fragility of the financial system typical of modern capitalist-globalized economies. An additional justification, due to Minsky, is that attempts at policy reform based on insufficient and inadequate knowledge of the underlying processes are likely to have limited success and may potentially cause even more harm. Unlike the modern mainstream of the economics profession, the reviewed non-mainstream works not only open up for the existence of endogenously occurring financial crises, but also rationalize them as a necessary product of the way modern capitalist-globalized economies operate. Indeed, crisis prevention should be concerned solely with avoidable events, which exogenously occurring financial crises — the only ones allowed by mainstream economic theory — apparently are not¹²⁴; the latter require improved crisis resolution mechanisms, to which, according to the reviewed post-crisis literature, corrective taxation can contribute only indirectly and to a limited degree. Some recognition of the merits of the non-mainstream theories explaining financial instability has been accorded in the post-crisis mainstream literature, which is a step in the right direction.

I then argued that the inefficient allocation, the expression of which was the recent crisis, is indicative of market failure, in the form of first and foremost the financial sector's exerting significant negative externalities on the rest of the economy. The latter are realized through systemic risk, the containment of which is a task of macroprudential policy. Drawing a parallel to their use to deal with environmental externalities, the suitability of corrective taxes and fees as macroprudential tools has been subject to discussion in the wider post-crisis debate on appropriate policy measures to rein in the excesses of the financial sector by aligning its incentives with the socially optimal ones. While there appears to be no consensus on how (dis)similar environmental and financial externalities are, there is agreement that certain features of the latter do not allow direct transfer of solutions. Nevertheless, there is role for corrective taxes – in theory, they can be used to address the various issues which contribute to the build-up of systemic risk — excessive risk-taking, debt bias, reliance on short-term debt to name a few — by forcing the financial sector to face the full costs of its actions. Further, they can be used to correct certain deplorable institutional arrangements — the implicit subsidy and the debt guarantee enjoyed by the financial sector, contributing to subsidized

¹²⁴ Strictly speaking, they are, but only by chance. Reliance on coincidences is not what one would expect from solid science, or policy makers. As far as the latter are concerned, however, it is a different matter whether reliance on them or chance produces lesser harm.

borrowing, wrong pricing of risk and an oversized industry — which are the source of moral hazard and hence are conducive to financial instability. Lastly, I have come with a simple proposal to use a fee as a means of reducing the scope of informational advantage the financial sector enjoys. In practice, however, things are neither so clear-cut, nor particularly simple: currently there are few implementable options available and there are neither so many more that look promising. The latter can probably be explained in part by the lack of better understanding of the interaction between the financial sector and the rest of the economy; this necessitates a major improvement of the set of models in economists' toolkits, a point agreed on in the post-crisis literature.

Some limitations of the approach have been identified. First, the presence of financial frictions means that only second-best options are available. Second, systemic institutions will require separate treatment, because empirical studies tend to indicate that there are both quantitative and qualitative aspects to such entities. There seems to be agreement in the post-crisis literature that a mixture of macroprudential approaches towards systemic institutions is the way to go. With the exception of Minsky, the reviewed literature elegantly skips the question of whether there is economic justification for the existence of such entities. Further, it accords, what would appear, marginal attention to what I have termed *systemicness bias* — the conscious attempts by financial institutions to become systemic in order that they enjoy the advantages this status confers.

Certain features of the particular tax design have been identified in the literature. To determine the tax base, the boundaries of "financial activities" will have to be defined more precisely. Closely related is the need to develop a measure of the externality exerted by the financial sector. Several such proposals have been put forward, all relying on the market mechanism to provide the necessary information and, defying experience from the recent crisis, in all papers dealing with the issue it is taken as given that the information will be the actual one. With a view to reflecting the contribution to externalities by each institution, tax rates will likely be differentiated and virtually certainly non-linear, many of them time-varying as well. The incidence of many of the proposed variants is unclear, which is problematic not just within the context of the current composition, but also for wider issues, such as efficiency and equity. International coordination is highly desirable to achieve maximum effect, but is not a must, provided effective measures against avoidance are available. All proposed taxes may potentially provide non-insignificant revenue.

Three types of taxes have been put forward: FTTs, FATs, and levies on financial institutions' liabilities. The first group have been dismissed for an array of reasons. Theoretical results seem to strongly indicate the ineptitude of FTTs to meet macroprudential goals, while empirical studies come to ambiguous conclusions. They may further have perverse, counter-intended results, which is welfare-reducing. Further, FTTs are dismissed for efficiency reasons, for instance their incidence is unclear, but a large fraction of it is virtually sure to fall outside the financial sector. Thus, too much ambiguity and situation-dependence make it hard to come with some general rules with wide applicability. Studies show that not even in one intuitive area of application — the repeal of the financial sector's VAT exemption — are FTTs the optimal way to implement such a policy. On the other hand FATs have been considered a viable option to realize Pareto-improved allocations. Their acceptance is rooted in efficiency reasons — they will fall on rents. There is also less ambiguity around their base — it will comprise the profits and wages of financial institutions. However, FATs are neither universally applicable, nor are without problems, for instance their effect on variable remuneration is unclear in general. The third group — levies on financial institutions' liabilities — are regarded as useful substitutes, complements and improvement over existing arrangements, such as the use of capital requirements. There is more or less unanimity that their base will comprise financial institutions' liabilities. Their incidence is not quite clear in general, but will depend on the market structure.

The overall conclusion that can be drawn from the reviewed literature is that there is scope for applying corrective taxes and fees in the work towards a financial sector which is more robust and serves its purposes as postulated by economic theory, and their use should be coordinated with

other measures to achieve optimal macroprudential results. Empirical research in many areas is required to find the best mixture for each particular area of application.

References

- Acharya, Viral V., Lasse H. Pedersen, Thomas Philippon & Matthew Richardson.** 2010. "A Tax on Systemic Risk". *Working Paper, first draft*. NY Stern
- Allen, Franklin, & Anthony M. Santomero.** 1999. "What Do Financial Intermediaries Do?". *Working Paper Series*, The Wharton School: 1-24
- Bean, Charles.** 2009. "The Great Moderation, the Great Panic and the Great Contraction". *Annual Congress of the European Economic Association*, Schumpeter Lecture, Barcelona, 25th August 2009
- Bernanke, Ben S., Mark Gertler & Simon Gilchrist.** 1999. "The Financial Accelerator in a Quantitative Business Cycle Framework". Chapter 21 in J. B. Taylor & M. Woodford (eds.) *Handbook of Macroeconomics, Vol. 1*: 1341-1393. Elsevier Science B.V.
- Brunnermeier, Markus K.** 2009. "Deciphering the Liquidity and Credit Crunch 2007-2008". *Journal of Economic Perspectives*, Volume 23, Number 1. Winter 2009: 77-100
- Buettner, Thiess & Katharina Erbe.** 2012. "Revenue and Welfare Effects of Financial Sector VAT Exemption". *TaxFACTs Schriftenreihe* Nr. 2012-02
- Cooper, George.** 2008. *The Origin of Financial Crises*. Petersfield: Harriman House Ltd.
- Cottarelli, Carlo.** 2010. "Fair and Substantial – Taxing the Financial Sector". In Stijn Claessens, Michael Keen & Ceyla Pazarbaşıoğlu (eds.) *Financial Sector Taxation: The IMF's Report to the G-20 and Background Material*: 74-77
- de Nicolò, Gianni.** 2010. "The Potential Impact of FSC and FAT on Bank Asset Growth, Risk, and Real Activity". In Stijn Claessens, Michael Keen & Ceyla Pazarbaşıoğlu (eds.) *Financial Sector Taxation: The IMF's Report to the G-20 and Background Material*: 188-191
- Diamond, Peter & James Mirrlees.** 1971. "Optimal Taxation and Public Production 1: Production Efficiency". *American Economic Review* 61(3): 8-27
- Dolls, Mathias, Clemens Fuest & Andreas Peichl.** 2012. "Automatic Stabilizers and Economic Crisis: US vs. Europe". *Journal of Public Economics*, Vol. 96(3): 279-294
- Epstein, Gerald A.** (ed.). 2005. *Financialization and the World Economy*. Cheltenham: Edward Elgar
- Financial Times.** 2007. "Citigroup chief stays bullish on buy-outs". www.ft.com/cms/s/0/80e2987a-2e50-11dc-821c-0000779fd2ac.html#axzz3Ay2NrJrL. 9th July 2007
- Fisher, Irving.** 1933. "The Debt-Deflation Theory of Great Depressions". *Econometrica*, vol. 1, no. 4: 337-357
- Friedman, Milton.** 1953. "The Case for Flexible Exchange Rates" in Milton Friedman (ed.), *Essays in Positive Economics*. Chicago: University of Chicago Press

- Galbraith, John Kenneth.** 1993. *A Short History of Financial Euphoria*. New York: Viking Penguin
- Galbraith, John Kenneth.** 1995. *Money: Whence It Came, Where It Went*, 2nd edition. London: Penguin Books
- Huang, Rocco & Lev Ratnovski.** 2009. "The Dark Side of Bank Wholesale Funding". *Journal of Financial Intermediation*, 20 (2011): 248-263
- Huizinga, Harry, Johannes Voget & Wolf Wagner.** 2011. "International Taxation and Cross-Border Banking". European Banking Center Discussion Paper No. 2011-015
- IMF.** 2010. "A Fair and Substantial Contribution by the Financial Sector". In Stijn Claessens, Michael Keen & Ceyla Pazarbaşıoğlu (eds.) *Financial Sector Taxation: The IMF's Report to the G-20 and Background Material*: 2-73
- Jeanne, Olivier & Anton Korinek.** 2010. "Managing Credit Booms and Busts: A Pigouvian Taxation Approach". *NBER Working Paper Series*. Working Paper 16377
- Keen, Michael & Ruud de Mooij.** 2012. "Debt, Taxes, and Banks". *IMF Working Paper* WP/12/48
- Keen, Michael, Russell Krelove & John Norregaard.** 2010. "The Financial Activities Tax". In Stijn Claessens, Michael Keen & Ceyla Pazarbaşıoğlu (eds.) *Financial Sector Taxation: The IMF's Report to the G-20 and Background Material*: 118-143
- Keen, Michael.** 2011. "Rethinking the Taxation of the Financial Sector". *CESifo Economic Studies*, Vol. 57, 1/2011
- Keen, Michael.** 2012. "New Thinking on the Taxation of the Financial Sector". *Norway Tax Forum*, 12th June 2012. (The provision of the slides from Mr. Keen's presentation to me by Vidar Christiansen is gratefully acknowledged.)
- Keynes, John Maynard.** 1964. *The General Theory of Employment Interest and Money*. London: Macmillan & Co Ltd
- Kocherlakota, Narayana.** 2010. "Taxing Risk and the Optimal Regulation of Financial Institutions". Economic Policy Paper 10-3. Federal Reserve Bank of Minneapolis
- Korinek, Anton.** 2011a. "Systemic Risk-Taking. Amplification Effects, Externalities, and Regulatory Responses". *ECB Working Paper Series*, No. 1345/June 2011: 3-41
- Korinek, Anton.** 2011b. "Systemic Risk-Taking: Amplification Effects, Externalities, and Regulatory Responses". Presentation at the *Conference on Systemic Risk and Data Issues*. Washington, DC, October 2011
- Laeven, Luc & Fabián Valencia.** 2010a. "How Costly are the 2007-2009 Systemic Banking Crises?". In Stijn Claessens, Michael Keen & Ceyla Pazarbaşıoğlu (eds.) *Financial Sector Taxation: The IMF's Report to the G-20 and Background Material*: 78-83
- Laeven, Luc & Fabián Valencia.** 2010b. "Resolution of Banking Crises: The Good, The Bad, and The Ugly". *IMF Working Paper* WP10/146

- Laibson, David.** 1997. "Golden Eggs and Hyperbolic Discounting". *Quarterly Journal of Economics*, 112: 443-477
- Lockwood, Ben.** 2010. "How should Financial Intermediation Services be Taxed?" *CESifo Working Paper No. 3226*
- Mager, Nathan H..** 1987. *The Kondratieff Waves*. New York: Praeger Publishers
- Matheson, Thornton.** 2011. "Taxing Financial Transactions: Issues and Evidence". *IMF Working Paper WP/11/54*
- Minsky, Hyman P.** 1982. *Inflation, Recession and Economic Policy*. Brighton: Wheatsheaf Books Ltd.
- Minsky, Hyman P.** 1986. *Stabilizing an Unstable Economy*. A Twentieth Century Fund Report. Haven and London: Yale University Press
- Mirrlees, James, Stuart Adam, Tim Besley, Richard Blundell, Stephen Bond, Robert Chote, Malcolm Gammie, Paul Johnson, Gareth Myles, James Poterba.** 2011. *The Mirrlees Review*
- NOU.** 1990. "Generell merverdiavgift på omsetning av tjenester". *Norges offentlige utredninger*, 1990:11 (Norwegian Public Reports. *General Value-added (Excise) Tax on Sale of Services*)
- NOU.** 2009. "Skatteparadis og utvikling". *Norges offentlige utredninger*, 2009:19 (Norwegian Public Reports. *Tax Havens and Development*)
- NOU.** 2011. "Bedre rustet mot finanskriser: Finanskriseutvalgets utredning". *Norges offentlige utredninger*, 2011:1 (Norwegian Public Reports. *Better Equipped against Financial Crises: The Financial Crisis Committee's Report*)
- Palley, Thomas I.** 2007. "Financialization: What is it and Why it Matters". *Political Economy Research Institute, University of Massachusetts Amherst, Working Paper Series Number 153*
- Reinhart, Carmen M. & Kenneth Rogoff.** 2009. *This Time Is Different: Eight Centuries of Financial Folly*. Princeton: Princeton University Press
- Rødseth, Asbjørn.** 2000. *Open Economy Macroeconomics*. Cambridge: Cambridge University Press
- Schjelderup, Guttorm.** 2011. "Sekretesjejurisdiksjoner, korrupsjon og økonomisk kriminalitet". SNF Arbeidsnotat Nr 33/11 (*Jurisdictions of Secrecy, Corruption and Economic Criminality*)
- Shaviro, Daniel N., Douglas A. Shackelford, Joel B. Slemrod.** 2011. "Taxation and the Financial Sector". *NYU Law and Economics Research Paper No.10-25*
- Shin, Hyun Song.** 2010. "Non-Core Liabilities Tax as a Tool for Prudential Regulation". Policy Memo
- Spahn, Paul Bernd.** 1995. "International Financial Flows and Transaction Taxes: Survey and Options". *IMF Working Paper WP/95/60*
- Stiglitz, Joseph.** 1989. "Using Tax Policy To Curb Speculative Short-Term Trading". *Journal of Financial Services Research*, 3: 101-115

Tobin, James. 1978. "A Proposal for International Monetary Reform". *Eastern Economic Journal*, Volume (Year): 4 (1978), Issue (Month): 3-4 (July-October): 153-159

Tylecote, Andrew. 1992. *The Long Wave In The World Economy: The Present crisis in historical perspective*. London and New York: Routledge

Ueda, Kenichi & Beatrice Weder di Mauro. 2010. "The Value of the Too-Big-to-Fail Subsidy to Financial Institutions". In Stijn Claessens, Michael Keen & Ceyla Pazarbaşıoğlu (eds.) *Financial Sector Taxation: The IMF's Report to the G-20 and Background Material*: 106-117

von Ehrlich, Maximilian & Doina Radulescu. 2012. "The Taxation of Bonuses and its Effect on Executive Compensation and Risk Taking – Evidence from the UK Experience". *CESifo Area Conference on Public Sector Economics*, 12th -14th April 2012, Munich

von Mises, Ludwig. 1953. *The Theory Of Money And Credit*. New Haven: Yale University Press

von Mises, Ludwig. 1996. *Human Action: A Treatise On Economics*, 4th edition. San Francisco: Fox & Wilkes

Watson, Joel. 2008. *Strategy: An Introduction to Game Theory*, 2nd edition. New York: Norton.

Weder di Mauro, Beatrice & Ulrich Klüh. 2010. "Reshaping Systemic Risk Regulation in Europe". In Stijn Claessens, Michael Keen & Ceyla Pazarbaşıoğlu (eds.) *Financial Sector Taxation: The IMF's Report to the G-20 and Background Material*: 84-95

Weder di Mauro, Beatrice. 2010a. "Quantitative Impact of Taxing or Regulating Systemic Risk". In Stijn Claessens, Michael Keen & Ceyla Pazarbaşıoğlu (eds.) *Financial Sector Taxation: The IMF's Report to the G-20 and Background Material*: 96-105

Weder di Mauro, Beatrice. 2010b. "Taxing Systemic Risk: Proposal for a Systemic Risk Levy and a Systemic Risk Fund". Presented at the Deutsche Bundesbank, 25th January 2010

Windolf, Paul. 2005. "Was ist Finanzmarkt-Kapitalismus?" in *Finanzmarkt-Kapitalismus. Kölner Zeitschrift für Soziologie und Sozialpsychologie, Sonderheft 45/2005*: 20-57 (What is Financial Market Capitalism?)

Web resources:

"Austrian School". http://en.wikipedia.org/wiki/Austrian_School [accessed 9th February 2011]

"Bank tax". http://en.wikipedia.org/wiki/Bank_tax [accessed 22nd February 2012]

"Credit Cycle". http://en.wikipedia.org/wiki/Credit_cycle [accessed 21st July 2011]

"Dot-com Bubble". http://en.wikipedia.org/wiki/Dot-com_bubble [accessed 31st October 2012]

"European Union financial transaction tax".

http://en.wikipedia.org/wiki/European_Union_financial_transaction_tax [accessed 22nd February 2012]

“Financial transaction tax”. http://en.wikipedia.org/wiki/Financial_transaction_tax [accessed 22nd February 2012]

“Financialization”. <http://en.wikipedia.org/wiki/Financialization> [accessed 9th June 2014]

“Hyman Minsky”. http://en.wikipedia.org/wiki/Hyman_Minsky [accessed 23rd May 2012]

“John Kenneth Galbraith”. http://en.wikipedia.org/wiki/John_Kenneth_Galbraith [accessed 23rd May 2012]

“Juglar cycle”. http://en.wikipedia.org/wiki/Juglar_cycle [accessed 21st July 2011]

“Kitchin cycle”. http://en.wikipedia.org/wiki/Kitchin_cycle [accessed 21st July 2011]

“Kondratiev wave”. http://en.wikipedia.org/wiki/Kondratiev_wave [accessed 21st July 2011]

“Kuznets swing”. http://en.wikipedia.org/wiki/Kuznets_swing [accessed 21st July 2011]

“Late-2000s financial crisis”. http://en.wikipedia.org/wiki/Financial_crisis_of-2007%E2%80%932010 [accessed 22nd February 2012]

“Macroprudential regulation”. http://en.wikipedia.org/wiki/Macroprudential_regulation [accessed 20th April 2013]

“Pigovian tax”. http://en.wikipedia.org/wiki/Pigovian_tax [accessed 21st March 2011]

“Ramsey problem”. http://en.wikipedia.org/wiki/Ramsey_problem [accessed 22nd February 2012]

“Reaction to the Tobin Tax”. http://en.wikipedia.org/wiki/Reaction_to_the_Tobin_Tax [accessed 22nd February 2012]

“Swedish financial transaction tax”. http://en.wikipedia.org/wiki/Swedish_financial_transaction_tax [accessed 22nd February 2012]

“Tobin tax”. http://en.wikipedia.org/wiki/Tobin_tax [accessed 21st March 2011]

“Valpskatt”. <http://sv.wikipedia.org/wiki/Valpskatt> [accessed 22nd February 2012]

“Causes of the late-2000s financial crisis”. http://en.wikipedia.org/wiki/Causes_of_the_late-2000s_financial_crisis [accessed 22nd February 2012]